

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:35:35 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatctcaggaattatcaaa.....tcagtcgaggaacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	419.6	83.8	1158	1 US-08-471-724-1	Sequence 1, Appli
2	419.6	83.8	1158	2 US-08-471-969-1	Sequence 1, Appli
3	419.6	83.8	1158	2 US-08-384-137-1	Sequence 1, Appli
4	419.6	83.8	1158	2 US-08-470-006A-1	Sequence 1, Appli
5	419.6	83.8	1158	3 US-08-691-563C-1	Sequence 1, Appli
6	419.6	83.8	1158	4 US-09-200-990-1	Sequence 1, Appli
7	419.6	83.8	1158	4 US-09-133-411-1	Sequence 1, Appli
8	419.6	83.8	2391	3 US-08-691-563C-57	Sequence 57, Appli
9	404.4	80.7	1577	3 US-08-691-563C-89	Sequence 89, Appli
10	271.8	54.3	2330	4 US-09-120-653D-4	Sequence 4, Appli
11	271.8	54.3	3910	4 US-09-120-653D-1	Sequence 1, Appli
12	212.6	42.4	297	1 US-08-471-724-2	Sequence 2, Appli
13	212.6	42.4	297	2 US-08-471-969-2	Sequence 2, Appli
14	212.6	42.4	297	2 US-08-384-137-2	Sequence 2, Appli
15	212.6	42.4	297	2 US-08-470-006A-2	Sequence 2, Appli
16	212.6	42.4	297	3 US-08-691-563C-2	Sequence 2, Appli
17	212.6	42.4	297	4 US-09-200-990-2	Sequence 2, Appli
18	212.6	42.4	297	4 US-09-133-411-2	Sequence 2, Appli
19	211.4	42.2	645	1 US-08-471-724-8	Sequence 8, Appli
20	211.4	42.2	645	2 US-08-471-969-8	Sequence 8, Appli
21	211.4	42.2	645	2 US-08-384-137-8	Sequence 8, Appli
22	211.4	42.2	645	2 US-08-470-006A-8	Sequence 8, Appli
23	211.4	42.2	645	3 US-08-691-563C-8	Sequence 8, Appli
24	211.4	42.2	645	4 US-09-200-990-8	Sequence 8, Appli
25	211.4	42.2	645	4 US-09-133-411-8	Sequence 8, Appli
26	128.8	25.7	299	3 US-08-691-563C-40	Sequence 40, Appli
27	116.4	23.2	2448	3 US-08-691-563C-53	Sequence 53, Appli

28	109.8	21.9	741	1 US-08-471-724-9	Sequence 9, Appli
29	109.8	21.9	741	2 US-08-471-969-9	Sequence 9, Appli
30	109.8	21.9	741	2 US-08-384-137-9	Sequence 9, Appli
31	109.8	21.9	741	2 US-08-470-006A-9	Sequence 9, Appli
32	109.8	21.9	741	3 US-08-691-563C-9	Sequence 9, Appli
33	109.8	21.9	741	4 US-09-200-990-9	Sequence 9, Appli
34	109.8	21.9	741	4 US-09-133-411-9	Sequence 9, Appli
35	109.8	21.9	2389	3 US-08-691-563C-52	Sequence 52, Appli
36	92.6	18.5	4480	4 US-09-167-322-12	Sequence 12, Appli
37	86.6	17.3	6363	2 US-08-929-967-6	Sequence 6, Appli
38	85.8	17.1	1140	4 US-09-603-185-5	Sequence 5, Appli
39	85.8	17.1	3612	4 US-09-265-013-3	Sequence 3, Appli
40	85.8	17.1	7308	4 US-09-011-745-3	Sequence 3, Appli
41	85.8	17.1	7308	4 US-09-011-745-4	Sequence 4, Appli
42	85.8	17.1	7616	4 US-09-011-745-2	Sequence 2, Appli
43	85.8	17.1	8202	1 US-08-258-420-13	Sequence 13, Appli
44	85.8	17.1	8332	3 US-08-850-961-1	Sequence 1, Appli
45	85.8	17.1	8332	4 US-09-479-776-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-471-724-1
Sequence 1, Application US/08471724
Patent No. 5800980
GENERAL INFORMATION:
APPLICANT: HERVE PERON
APPLICANT: FRANCOIS MAILLET
APPLICANT: BERNARD MANDRAND
APPLICANT: FREDERIC BEDIN
APPLICANT: FREDERIC BESME
TITLE OF INVENTION: MERV1 VIRUS AND MERV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOLY
NUMBER OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
TITLE OF INVENTION: 38
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,724
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-724-1
Query Match 83.8%; Score 419.6; DB 1; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;
CY 1 CAAGATCTCAGGATTATCAATGAGAGCTGTGTTCTCTATGACCAAGCTGTACTAGCCCT 60

Db 61 CAAGAACTCAGGATTATCAATGAGGCTGTGTCCCTCTATACCCAGCTGTACTTAACCT 120
Qy 61 TATACTGCTTTTCCCAATATACGAGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 120
Db 121 TATACAGTCTTTTCCCAATATACGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 180
Qy 121 GATGCTCTTCTGTGATCCCTGTACATCTGTGATCTCAATCTGTGCTTGTGAAGAT 180
Db 181 GATGCTCTTCTGTGATCCCTGTGATCTGTGATCTCAATCTGTGCTTGTGAAGAT 240
Qy 181 ACTTCAAAACCAATCTCACTCACTGAGTATTTTACCCCAAGGTTTCAAGGATAGT 240
Db 241 CTTTGAACCCAAAGCTCACTCACTGAGTATTTTACCCCAAGGTTTCAAGGATAGC 300
Qy 241 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCATTCCTCACTGAGACA-- 298
Db 301 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCATTCCTCACTGAGACT 360
Qy 239 CTTGTCTTGGGTAGGTGATGATTTTCTTTGGCCGCCATTCAGAAACCTTGTGCAT 358
Db 361 CTTGTCTTGGGTAGGTGATGATTTTCTTTGGCCGCCATTCAGAAACCTTGTGCAT 420
Qy 359 CAAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGCTCACTGTTTCCAAACCA 418
Db 421 CAAGCCACCAAGAACTTTAACTTTCTCTACCTGTGCTCACTGTTTCCAAACCA 480
Qy 419 AAGGCTCACTGCTGCTCAAGAG---GTTACTTAAGGCTTAAATATTCAAAGGCACC 474
Db 481 AAGGCTGCTGCTGCTCAAGAGATTAGATCTTAAGGCTTAAATATTCAAAGGCACC 540
Qy 475 AAGGCTCTGAGTGAAGACATCCAG 501
Db 541 AAGGCTCTGAGTGAAGACATCCAG 567

RESULT 2
US-08-471-969-1
; Sequence 1, Application US/08471969
; Patent No. 5871745

; GENERAL INFORMATION:
; APPLICANT: Hervé PERRON
; APPLICANT: Francois MALLET
; APPLICANT: Bernard MANDRAND
; APPLICANT: Frederic BESME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESS: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,969
; FILING DATE: June 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36055A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
; LENGTH: 1158 bases
; TYPE: nucleotide
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-471-969-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2,5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

Qy 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTCTCTATAGCCAGCTGTACTTAACCTT 60
Db 61 CAAGAACTCAGGATTATCAATGAGGCTGTGTCTCTATACCCAGCTGTACTTAACCTT 120
Qy 61 TATACTGCTTTTCCCAATATACGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 120
Db 121 TATACAGTCTTTTCCCAATATACGAGAGAGAGTGGTTTACAGTCTGTGACCTTCAAG 180
Qy 121 GATGCTCTTCTGTGATCCCTGTACATCTGTGATCTCAATCTGTGCTTGTGAAGAT 180
Db 181 GATGCTCTTCTGTGATCCCTGTACATCTGTGATCTCAATCTGTGCTTGTGAAGAT 240
Qy 181 ACTTCAAAACCAATCTCACTCACTGAGTATTTTACCCCAAGGTTTCAAGGATAGT 240
Db 241 CTTTGAACCCAAAGCTCACTCACTGAGTATTTTACCCCAAGGTTTCAAGGATAGC 300
Qy 241 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCATTCCTCACTGAGACA-- 298
Db 301 CCCCATCTATTGGGCGAGGATTTAGCCCAAGCTTGAAGCATTCCTCACTGAGACT 360
Qy 239 CTTGTCTTGGGTAGGTGATGATTTTCTTTGGCCGCCATTCAGAAACCTTGTGCAT 358
Db 361 CTTGTCTTGGGTAGGTGATGATTTTCTTTGGCCGCCATTCAGAAACCTTGTGCAT 420
Qy 359 CAAGCCACCAAGCGCTCTTCAATTTCTGCTACCTGTGCTCACTGTTTCCAAACCA 418
Db 421 CAAGCCACCAAGAACTTTAACTTTCTCTACCTGTGCTCACTGTTTCCAAACCA 480
Qy 419 AAGGCTCACTGCTGCTCAAGAG---GTTACTTAAGGCTTAAATATTCAAAGGCACC 474
Db 481 AAGGCTGCTGCTGCTCAAGAGATTAGATCTTAAGGCTTAAATATTCAAAGGCACC 540
Qy 475 AAGGCTCTGAGTGAAGACATCCAG 501
Db 541 AAGGCTCTGAGTGAAGACATCCAG 567

RESULT 3
US-08-384-137-1
; Sequence 1, Application US/08384137
; Patent No. 5871996

; GENERAL INFORMATION:
; APPLICANT: Hervé PERRON
; APPLICANT: Francois MALLET
; APPLICANT: Bernard MANDRAND
; APPLICANT: Frederic BESME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESS: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,137
; FILING DATE: June 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36055A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384.137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-384-137-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140; Indels 6; Gaps 2;
Matches 466; Conservative 0; Mismatches 35;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCT 180
DB 181 GATGCTCTCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCT 240
QY 181 ACTTCAAAACCAACATCTCACTCACTGACTATTTTACCCTCAAGGCTTCAAGGATAGT 240
DB 241 CTTTGAACCCCAAGCTCTCACTCACTGACTGTTTACCCTCAAGGCTTCAAGGATAGC 300
QY 241 CCCCATTATTTGGCCAGGATTAAGCCCAAGCTTGAATCTCAATCCTGAGACACT 298
DB 301 CCCCATTATTTGGCCAGGATTAAGCCCAAGCTTGAATCTCAATCCTGAGACACT 360
QY 299 CTTGCTCTTGGTGGTGGATGATTTACTTTGGCCGCCCATTTCAAGAACCTTGGCCAT 358
DB 361 CTTGCTCTTGGTGGTGGATGATTTACTTTGGCCGCCCATTTCAAGAACCTTGGCCAT 420
QY 359 CAAGCCACCCAGGCTCTTCAATTTCTCGCTACCTGCTAGGCTACATGAGTTTCAAAACA 418
DB 421 CAAGCCACCCAGGATTAAGCTTTACTTCTCACTACCTGCTAGGCTTCAAGTTTCAAAACA 480
QY 419 AAGGCTCAACTCTGCTCAACAGAG----GTACTTAGGGCTAAATTAATCCAAAGGCAAC 474
DB 481 AAGGCTGGCTCTGCTCAACAGAGATTAGATCTAAAGGCTAAATTAATCCAAAGGCAAC 540
QY 475 AAGGCTCTCACTGAGAGAACATCCAG 501
DB 541 AAGGCTCTCACTGAGAGAACATCCAG 567

RESULT 4
US-08-470-006A-1
Sequence 1, Application US/08470006A
GENERAL INFORMATION:
APPLICANT: Herve PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESSEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
TITLE OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470.006A
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-470-006A-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140; Indels 6; Gaps 2;
Matches 466; Conservative 0; Mismatches 35;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCT 180
DB 181 GATGCTCTCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCT 240
QY 181 ACTTCAAAACCAACATCTCACTCACTGACTATTTTACCCTCAAGGCTTCAAGGATAGT 240
DB 241 CTTTGAACCCCAAGCTCTCACTCACTGACTGTTTACCCTCAAGGCTTCAAGGATAGC 300
QY 241 CCCCATTATTTGGCCAGGATTAAGCCCAAGCTTGAATCTCAATCCTGAGACACT 298
DB 301 CCCCATTATTTGGCCAGGATTAAGCCCAAGCTTGAATCTCAATCCTGAGACACT 360
QY 299 CTTGCTCTTGGTGGTGGATGATTTACTTTGGCCGCCCATTTCAAGAACCTTGGCCAT 358
DB 361 CTTGCTCTTGGTGGTGGATGATTTACTTTGGCCGCCCATTTCAAGAACCTTGGCCAT 420
QY 359 CAAGCCACCCAGGCTCTTCAATTTCTCGCTACCTGCTAGGCTACATGAGTTTCAAAACA 418
DB 421 CAAGCCACCCAGGATTAAGCTTTACTTCTCACTACCTGCTAGGCTTCAAGTTTCAAAACA 480
QY 419 AAGGCTCAACTCTGCTCAACAGAG----GTACTTAGGGCTAAATTAATCCAAAGGCAAC 474
DB 481 AAGGCTGGCTCTGCTCAACAGAGATTAGATCTAAAGGCTAAATTAATCCAAAGGCAAC 540
QY 475 AAGGCTCTCACTGAGAGAACATCCAG 501
DB 541 AAGGCTCTCACTGAGAGAACATCCAG 567

RESULT 5

US-08-691-563C-1

Sequence 1, Application US/08691563C

Patent No. 6001987

GENERAL INFORMATION:

APPLICANT: HERVE PERRON

APPLICANT: FREDERIC BESEME

APPLICANT: FREDERIC BESIM

APPLICANT: GLAUCIA PARANHOS-BACCALA

APPLICANT: FLORENCE KOMURIAN-PRADEL

APPLICANT: COLETTE JOLIVET

APPLICANT: BERNARD MANDRAND

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 92

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/691,563C

FILING DATE: 02-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 38588

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 base pairs

TYPE: nucleotide

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-691-563C-1

Query Match 83.8%; Score 419.6; DB 3; Length 1158;

Best Local Similarity 91.9%; Pred. No. 2.5e-140;

Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACTAGCCCT 60
DB 61 CAAGAAGCTCAGGATTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACTAGCCCT 120
QY 61 TATATCTGCTTTCCCAATATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTTTCTTGATCCCTGTAGTCTCTGATCTCAATTTCTTTGTTCCCTTTGAAGAT 180
DB 181 GATGCTTTCTTGATCCCTGTAGTCTCTGATCTCAATTTCTTTGTTCCCTTTGAAGAT 240
QY 181 ACTTCAAAACCAACATCTCACTACCTGAGCTATTTTATCCCAAGGCTTCAAGGATAGT 240
DB 241 CTTTGAACCCCAACGCTCACTACCTGAGCTGTTTATCCCAAGGCTTCAAGGATAGC 300
QY 241 CCCCATCTATTTGGCAGGAGCTTAAAGCCCAAGCTTGAAGCATCTCTCACTTGAACA-- 298
DB 301 CCCCATCTATTTGGCAGGAGCTTAAAGCCCAAGCTTGAAGCATCTCTCACTTGAACA 360

QY 299 CTTCGCTTCGAGTAGTGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGCCAT 358
DB 361 CTTCGCTTCGAGTAGTGATGATTTACTTTAGTCCGCCGCTCAGAAACCTTGCCAT 420
QY 359 CAAGCCACCCCAAGGCTTTCAATTTCTGCTACCTGTGGCTTCATGTTCCAAACCA 418
DB 421 CAAGCCACCCCAAGGCTTTCAATTTCTGCTACCTGTGGCTTCATGTTCCAAACCA 480
QY 419 AAGGCTCACTGCTCTACAGCAGCAG----GTTACTAGGGCTTAAATTTCAAGAGCACC 474
DB 481 AAGGCTCGGCTCTGCTCTACAGAGATTTAGATCTAAGGGCTTAAATTTCAAGAGCACC 540
QY 475 AGGCGCTCTAGTAGAGAACATCTCAG 501
DB 541 AGGCGCTCTAGTAGAGAACATCTCAG 567

RESULT 6

US-09-200-990-1

Sequence 1, Application US/09200990

Patent No. 6184025

GENERAL INFORMATION:

APPLICANT: HERVE PERRON

APPLICANT: FRANCOIS MALLET

APPLICANT: BERNARD MANDRAND

APPLICANT: FREDERIC BESIME

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/200,990

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/471,969

FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 36055A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 bases

TYPE: nucleotide

STRANDEDNESS: single-stranded

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-200-990-1

Query Match 83.8%; Score 419.6; DB 4; Length 1158;

Best Local Similarity 91.9%; Pred. No. 2.5e-140;

Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACTAGCCCT 60
DB 61 CAAGAAGCTCAGGATTATCATGAGCTGTTGTTCTCTATACCGACGCTGTACTAGCCCT 120

```

QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGGTTTACAGTCTGAGCCTTCAG 120
DB 121 TATACAGTGTCTTCCCAATATACAGAGAGAGAGTGTGTTTACAGTCTGAGCCTTCAG 180
QY 121 GATGCTTCTTCTGATCCCTGATACATCTGACTCTCAATTCTTGTGCTTTGAGAGAT 180
DB 181 GATGCTTCTTCTGATCCCTGATACATCTGACTCTCAATTCTTGTGCTTTGAGAGAT 240
QY 181 ACTTCAAAACCAACATCTCACTACCTGACTATTTTACCACCAAGGTTTCAGGATAGT 240
DB 241 CTTTGAACCCCAACCTCTCACTACCTGACTGTTTACCACCAAGGTTTCAGGATAGC 300
QY 241 CCCCATCTATTGCGCAGGACATTTAGCCCAAGACTTGAACCAATCTCATCTGACAC 298
DB 301 CCCCATCTATTGCGCAGGACATTTAGCCCAAGACTTGAACCAATCTCATCTGACAC 360
QY 299 CTTGCTCTGCTGATGATGATTTTATTTTGGCCGCTTCAAGAACTTGTGCTCAT 358
DB 361 CTTGCTCTGCTGATGATGATTTTATTTTGGCCGCTTCAAGAACTTGTGCTCAT 420
QY 359 CAAGCACCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGTTTCCAAACCA 418
DB 421 CAAGCACCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGTTTCCAAACCA 480
QY 419 AAGGCTCACTCTGCTCACAGCAG---GTTACTTAGGGCTAAATATTCACAAAGGCACC 474
DB 481 AAGGCTGGGCTCTGCTCACAGAGATTAGATTACTTAGGGCTAAATATTCACAAAGGCACC 540
QY 475 AAGGCTCTCACTGAGGAACACATCCAG 501
DB 541 AAGGCTCTCACTGAGGAACATGATCCAG 567

```

RESULT 7

```

US-09-133-411-1
; Sequence 1, Application US/09133411
; Patent No. 6342383
; GENERAL INFORMATION:
; APPLICANT: Hervé PERRON
; APPLICANT: Francois MALLEF
; APPLICANT: Bernard MANDRAND
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BESEME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPRO
; TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/133,411
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/384,137
; FILING DATE: February 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36055
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1158 bases
; TYPE: nucleotide
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-133-411-1

```

```

Query Match 83.8%; Score 419.6; DB 4; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

```

```

QY 1 CAAGATCTCAGAGATTATCAATAGAGCTGTGTTCTCTATAGCCAGCTGTACCTTAC 60
DB 61 CAAGATCTCAGAGATTATCAATAGAGCTGTGTTCTCTATAGCCAGCTGTACCTTAC 120
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGGTTTACAGTCTGAGCCTTCAG 120
DB 121 TATACAGTGTCTTCCCAATATACAGAGAGAGAGTGTGTTTACAGTCTGAGCCTTCAG 180
QY 121 GATGCTTCTTCTGATCCCTGATACATCTGACTCTCAATTCTTGTGCTTTGAGAGAT 180
DB 181 GATGCTTCTTCTGATCCCTGATACATCTGACTCTCAATTCTTGTGCTTTGAGAGAT 240
QY 181 ACTTCAAAACCAACATCTCACTACCTGAGACTATTTTACCACCAAGGTTTCAGGATAGT 240
DB 241 CTTTGAACCCCAACCTCTCACTACCTGAGACTGTTTACCACCAAGGTTTCAGGATAGC 300
QY 241 CCCCATCTATTGCGCAGGACATTTAGCCCAAGACTTGAACCAATCTCATCTGACAC 298
DB 301 CCCCATCTATTGCGCAGGACATTTAGCCCAAGACTTGAACCAATCTCATCTGACAC 360
QY 299 CTTGCTCTGCTGATGATGATTTTATTTTGGCCGCTTCAAGAACTTGTGCTCAT 358
DB 361 CTTGCTCTGCTGATGATGATTTTATTTTGGCCGCTTCAAGAACTTGTGCTCAT 420
QY 359 CAAGCACCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGTTTCCAAACCA 418
DB 421 CAAGCACCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGTTTCCAAACCA 480
QY 419 AAGGCTCACTCTGCTCACAGCAG---GTTACTTAGGGCTAAATATTCACAAAGGCACC 474
DB 481 AAGGCTGGGCTCTGCTCACAGAGATTAGATTACTTAGGGCTAAATATTCACAAAGGCACC 540
QY 475 AAGGCTCTCACTGAGGAACACATCCAG 501
DB 541 AAGGCTCTCACTGAGGAACATGATCCAG 567

```

RESULT 8

```

US-08-691-563C-57
; Sequence 57, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: Hervé PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Frederic BESEME
; APPLICANT: Glauca PARANHOS-BACCALA
; APPLICANT: Florence KOMORIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTI
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-691-563C-57

Query Match 83.8%; Score 419.6; DB 3; Length 2391;
Best Local Similarity 91.9%; Pred. No. 3.8e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGTATATCAATGAGCGTGTGTTCTCTATATGCGAGCTGTACCTAGCCCT 60
DB 646 CAAGACTCAGAGTATATCAATGAGCGTGTGTTCTCTATATGCGAGCTGTACCTAGCCCT 705
QY 61 TATACTGCTTTCCCAATAACAGAGGAGAGGAGTGTATACAGTCTGTGACCTTCAAG 120
DB 706 TATACAGGCTTTCCCAATAACAGAGGAGAGTGTATACAGTCTGTGACCTTCAAG 765
QY 121 GATGCTTCTTCTGCATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
DB 766 GATGCTTCTTCTGCATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 825
QY 181 ACTTGAACCCCAACATCTCACTACCTGAGTATTTTACCCCAAGGTTGAGGATAGT 240
DB 826 CTTTGAACCCCAACATCTCACTACCTGAGTATTTTACCCCAAGGTTGAGGATAGT 885
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATTGAGCCATCTCTATACCTTGACA-- 298
DB 886 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATTGAGCCATCTCTATACCTTGACA-- 945
QY 299 CTTGTCTTGGTAGTGTGATTTACTTTGGCCGCCCATTTCAAGAACTTTGGCCAT 358
DB 946 CTTGTCTTGGTAGTGTGATTTACTTTGGCCGCCCATTTCAAGAACTTTGGCCAT 1005
QY 359 CAAGCACCCCAAGCGCTCTTCAATTTCCGCTACCTGTGAGTACATGGTTTCCAAACA 418
DB 1006 CAAGCACCCCAAGCGCTCTTCAATTTCCGCTACCTGTGAGTACATGGTTTCCAAACA 1065
QY 419 AAGGCTCACTCTGCTCAAGCAG--GTTACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 1066 AAGGCTCGGCTGTGCTCAAGCAGATTAGATCTAAGGGCTAAATTTATCCAAAGGACAC 1125
QY 475 AAGGCTCTCAAGTAGGAGAACATCCAG 501
DB 1126 AAGGCTCTCAAGTAGGAGAACATCCAG 1152

RESULT 9

US-08-691-563C-89
Sequence 89, Application US/08691563C
Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: HERVE PERRON
APPLICANT: FREDERIC BESEME
APPLICANT: FREDERIC BEDIN
APPLICANT: GAUCIA PARANHOS-BACCALA

APPLICANT: Florence KOMIRIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-691-563C-89

Query Match 80.7%; Score 404.4; DB 3; Length 1577;
Best Local Similarity 89.1%; Pred. No. 8.3e-135;
Matches 449; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

QY 1 CAAGATCTCAGAGTATATCAATGAGCGTGTGTTCTCTATATGCGAGCTGTACCTAGCCCT 60
DB 998 CAAGATCTCAGAGTATATCAATGAGCGTGTGTTCTCTATATGCGAGCTGTACCTAGCCCT 1057
QY 61 TATACTGCTTTCCCAATAACAGAGGAGAGTGTATACAGTCTGTGACCTTCAAG 120
DB 1058 TATACAGGCTTTCCCAATAACAGAGGAGAGTGTATACAGTCTGTGACCTTCAAG 1117
QY 121 GATGCTTCTTCTGCATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
DB 1118 GATGCTTCTTCTGCATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 1177
QY 181 ACTTGAACCCCAACATCTCACTACCTGAGTATTTTACCCCAAGGTTGAGGATAGT 240
DB 1178 CTTTGAACCCCAACATCTCACTACCTGAGTATTTTACCCCAAGGTTGAGGATAGT 1237
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATTGAGCCATCTCTATACCTTGACA-- 300
DB 1238 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATTGAGCCATCTCTATACCTTGACA-- 1297
QY 301 TGTCTTGGTAGTGTGATTTACTTTGGCCGCCCATTTCAAGAACTTTGGCCATCA 360
DB 1298 TGTCTTGGTAGTGTGATTTACTTTGGCCGCCCATTTCAAGAACTTTGGCCATCA 1357
QY 361 AGCCACCCCAAGCGCTTTCAATTTCTGCTACCTGTGAGTACATGGTTTCCAAAGCAAA 420
DB 1358 AGCCACCCCAAGCGCTTTCAATTTCTGCTACCTGTGAGTACATGGTTTCCAAAGCAAA 1417
QY 421 GGCTCACTGCTCAGCAGCGGT---TACTTAGGGCTAAATTTATCCAAAGGACACG 476
DB 1418 GGCTCACTGCTCAGCAGCGGTAAATTTACTTAGGGTTAAATTTATCCAAAGGACACG 1477

QY 477 GGCCCTCAGTGAAGAACATCCCA 500
Db 1478 GGCCCTCTGTGAGGAATGATATCCA 1501

RESULT 10

US-09-120-653D-4
; Sequence 4, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 4
; LENGTH: 2330
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-4

Query Match 54.3%; Score 271.8; DB 4; Length 2330;
Best Local Similarity 73.8%; Pred. No. 3.7e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTAATGATGAGCTGTGTTCTCTATATGCCAGCTGTATAGCCCT 60
Db 624 CAAGATCTTACATCAATCAATGACGAGTAATCTCTATATCCAGTTGATCCCAACCC 683
QY 61 TATATCTGCTTTCCCAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 664 TATATCTGCTTTCTCAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 743
QY 121 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAATCTTGTGGCTTTGAAGAT 180
Db 744 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAAGCTCTTGTGGCTTTGAAGAT 803
QY 181 ACTTCAAAACCAATCTCAATCTCACTGACTATTTTAAACCCAAAGGTTCAAGGAT 240
Db 804 CCAACAAACCAATCTCAATCTCACTGACTATTTTAAACCCAAAGGTTTAAAGGAT 863
QY 241 CCCCATCTATTTGGGAGGAGATTAAGCCCAAGCTTGAGCCATCTCTCA--TACCTGACA 298
Db 864 CCAATCTGTTGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 923
QY 299 CTGTCCTTCGTAAGTGAATTTAATTTTGGCGCCCATTTAGAAACCTTGGCCAT 358
Db 924 CTGTCCTTCATATGATGATGATTTAATTTTGGCTTACCAAGTTAGAAAGCTTGGCCAG 983
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTCGTACCTGTGCTGATGATGATGATGATGAT 418
Db 984 CAGGCTACTAGATCTCTTGAACCTTCAAGCTATCAAGGAGTCAAGGAGTCAAGGAG 1043
QY 419 AAGGCTCAACTCTGCTACAGAGT---TACTTAGGGCTTAAATTTTCCAAAGGAGC 474
Db 1044 AAAGCCCACTTTGCTCAAGAGGATTAATTTTAAATTTTAAATTTTAAATTTTAAAT 1103
QY 475 AGGCTCTCAGTGAAGAACATCCAG 501
Db 1104 AGGCTCTCAGTGAAGAACATCCAG 1130

RESULT 11

US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-WON
; APPLICANT: JUN, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

Query Match 54.3%; Score 271.8; DB 4; Length 3910;
Best Local Similarity 73.8%; Pred. No. 5e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTTAATGATGAGCTGTGTTCTCTATATGCCAGCTGTATAGCCCT 60
Db 1020 CAAGATCTTACATCAATCAATGACGAGTAATCTCTATATCCAGTTGATCCCAACCC 1079
QY 61 TATATCTGCTTTCCCAATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 1080 TATATCTGCTTTCTCAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1139
QY 121 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAATCTTGTGGCTTTGAAGAT 180
Db 1140 GATGCTCTCTCTGATCCCTGTACCTGATCTGATCTCAAGCTCTTGTGGCTTTGAAGAT 1199
QY 181 ACTTCAAAACCAATCTCAATCTCACTGACTATTTTAAACCCAAAGGTTCAAGGAT 240
Db 1200 CCAACAAACCAATCTCAATCTCACTGACTATTTTAAACCCAAAGGTTTAAAGGAT 1259
QY 241 CCCCATCTATTTGGGAGGAGATTAAGCCCAAGCTTGAGCCATCTCTCA--TACCTGACA 298
Db 1260 CCAATCTGTTGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1319
QY 299 CTGTCCTTCGTAAGTGAATTTAATTTTGGCGCCCATTTAGAAACCTTGGCCAT 358
Db 1320 CTGTCCTTCATATGATGATGATTTAATTTTGGCTTACCAAGTTAGAAAGCTTGGCCAG 1379
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTCGTACCTGTGCTGATGATGATGATGATGAT 418
Db 1380 CAGGCTACTAGATCTCTTGAACCTTCAAGCTATCAAGGAGTCAAGGAGTCAAGGAG 1439
QY 419 AAGGCTCAACTCTGCTACAGAGT---TACTTAGGGCTTAAATTTTCCAAAGGAGC 474
Db 1440 AAAGCCCACTTTGCTCAAGAGGATTAATTTTAAATTTTAAATTTTAAATTTTAAAT 1499
QY 475 AGGCTCTCAGTGAAGAACATCCAG 501
Db 1500 AGGCTCTCAGTGAAGAACATCCAG 1526

RESULT 12

US-08-471-724-2
; Sequence 2, Application US/08471724
; Patent No. 580980
; GENERAL INFORMATION:
; APPLICANT: HEVE PERRON
; APPLICANT: FRANCOIS MALLE
; APPLICANT: BERNARD MANDRAND

APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,724
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-2787
TELEFAX: 703-836-6400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-724-2

Query Match 42.4%; Score 212.6; DB 1; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGATTATCATATGAGCTGTGTTCTCTATATACCGAGCTGTACTAGCCCT 60
DB 61 CAAGACTCAGATTATCATATGAGCTGTGTTCTCTATATACCGAGCTGTACTAGCCCT 120
QY 61 TATACAGTCTTCCCAATATCCAGAGAGAGAGTGTATTACAGTCTGTGACCTTAG 120
DB 121 TATACAGTCTTCCCAATATCCAGAGAGAGAGTGTATTACAGTCTGTGACCTTAG 180
QY 121 GATGCTTTCTTCTGCATCCCTGTACATCTGTCAATCTTGTGCTTTGAAGAT 180
DB 181 GATGCTTTTCTGCATCCCTGTACATCTGTCAATCTTGTGCTTTGAAGAT 240
QY 181 ACTTGAACCCCAATCTCACTCACTGAGCTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCCCAATCTCACTCACTGAGCTATTTTACCCCAAGGTTCAAGG 295

RESULT 13
US-08-471-969-2
Sequence 2, Application US/08471969
Patent No. 5871745
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,969
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-969-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGATTATCATATGAGCTGTGTTCTCTATATACCGAGCTGTACTAGCCCT 60
DB 61 CAAGACTCAGATTATCATATGAGCTGTGTTCTCTATATACCGAGCTGTACTAGCCCT 120
QY 61 TATACAGTCTTCCCAATATCCAGAGAGAGAGTGTATTACAGTCTGTGACCTTAG 120
DB 121 TATACAGTCTTCCCAATATCCAGAGAGAGAGTGTATTACAGTCTGTGACCTTAG 180
QY 121 GATGCTTTCTTCTGCATCCCTGTACATCTGTCAATCTTGTGCTTTGAAGAT 180
DB 181 GATGCTTTTCTGCATCCCTGTACATCTGTCAATCTTGTGCTTTGAAGAT 240
QY 181 ACTTGAACCCCAATCTCACTCACTGAGCTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCCCAATCTCACTCACTGAGCTATTTTACCCCAAGGTTCAAGG 295

RESULT 14
US-08-384-137-2
Sequence 2, Application US/08384137
Patent No. 5871996
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-384-137-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCGCAGGATTATCAATGAGGCTGTGTCCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGAACTCAGGATTATCAATGAGGCTGTGTCCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACTGCTTTCCCAATATCAGAGAGCAGAGTGTTTACAGTCTGAGACCTTCAAG 120
DB 121 TATACAGTGCTTTCCCAATATCAGAGAGCAGAGTGTTTACAGTCTGAGACCTTCAAG 180
QY 121 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
DB 181 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 240
QY 181 ACTTCAACCAACATCTCACTCACTGAGCTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAAGCTCACTCACTGAGCTGTTTACCCCAAGGTTCAAGG 295

RESULT 15

US-08-470-006A-2
Sequence 2, Application US/08470006A
Patent No. 5962217

GENERAL INFORMATION:

APPLICANT: Hervé PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
TITLE OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,006A
FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-470-006A-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCGCAGGATTATCAATGAGGCTGTGTCCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGAACTCAGGATTATCAATGAGGCTGTGTCCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACTGCTTTCCCAATATCAGAGAGCAGAGTGTTTACAGTCTGAGACCTTCAAG 120
DB 121 TATACAGTGCTTTCCCAATATCAGAGAGCAGAGTGTTTACAGTCTGAGACCTTCAAG 180
QY 121 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
DB 181 GATGCTTTCTTGCAATCCCTGTACATCTGACTCTCAATTTCTGTTGCTTTGAAGAT 240
QY 181 ACTTCAACCAACATCTCACTCACTGAGCTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCAAGCTCACTCACTGAGCTGTTTACCCCAAGGTTCAAGG 295

Search completed: April 17, 2003, 07:38:02
Job time: 39.6897 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:57:20 ; Search time 123.415 Seconds
(without alignments)
4077.979 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501

Sequence: 1 caagatcgcagatcatca.....tcagtgcagacacacacag 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.8	86.2	1600	7	US-08-979-847-207
2	425.6	85.0	1600	7	US-08-979-847-206
3	422.2	84.3	2304	7	US-08-979-847-87
4	419.6	83.8	1158	7	US-08-979-847-1
5	419.6	83.8	1597	7	US-08-979-847-205
6	419.6	83.8	2365	7	US-08-979-847-88
7	419.6	83.8	2391	7	US-08-979-847-53
8	404.4	80.7	1577	7	US-08-979-847-83
9	402.2	80.3	568	10	US-09-864-761-16537
10	236.6	47.2	557	10	US-09-864-761-11881
11	231.6	46.2	449	10	US-09-864-761-11921
12	221.6	44.2	467	10	US-09-864-761-1194
13	212.6	42.4	297	7	US-08-979-847-2
14	212	42.3	540	10	US-09-864-761-15266
15	211.4	42.2	645	7	US-08-979-847-8
16	182.6	36.4	438	7	US-08-979-847-129
17	182.6	36.4	438	7	US-08-979-847-130
18	182.2	36.4	438	7	US-08-979-847-128
19	182.2	36.4	438	7	US-08-979-847-131

20	181.8	36.3	438	7	US-08-979-847-201	Sequence 201, App
21	165.6	33.1	944	10	US-09-864-761-14911	Sequence 14911, A
22	146.6	29.3	429	7	US-08-979-847-138	Sequence 138, App
23	145.4	29.0	429	7	US-08-979-847-135	Sequence 135, App
24	145	28.9	429	7	US-08-979-847-136	Sequence 136, App
25	145	28.9	429	7	US-08-979-847-137	Sequence 137, App
26	137.2	27.4	150	7	US-08-979-847-36	Sequence 46, App1
27	116.4	23.2	2448	7	US-08-979-847-49	Sequence 49, App1
28	114.2	22.8	191	10	US-09-864-761-19702	Sequence 19702, A
29	109.8	21.9	741	7	US-08-979-847-9	Sequence 9, App1
30	109.8	21.9	2389	7	US-08-979-847-48	Sequence 48, App1
31	107.6	21.5	133	10	US-09-864-761-33021	Sequence 33021, A
32	104.4	20.8	431	10	US-09-864-761-28100	Sequence 28100, A
33	102.2	20.4	181	10	US-09-864-761-17958	Sequence 17958, A
34	92.6	18.5	447	10	US-09-864-761-11512	Sequence 11512, A
35	92.4	18.4	326014	10	US-09-731-231A-3	Sequence 3, App1
36	89	17.8	157	10	US-09-864-761-31788	Sequence 31788, A
37	85.8	17.1	2151	10	US-09-845-157-1	Sequence 1, App1
38	85.8	17.1	8332	10	US-09-006-298-1	Sequence 1, App1
39	82.8	16.5	143068	10	US-09-967-768A-316	Sequence 316, App
40	77.8	15.5	8323	10	US-09-970-597-2	Sequence 2, App1
41	76.8	15.3	183	10	US-09-864-761-28561	Sequence 28561, A
42	76	15.2	685	10	US-09-864-761-31430	Sequence 31430, A
43	70.8	14.1	1436	10	US-09-864-761-17414	Sequence 17414, A
44	69.8	13.9	85	7	US-08-979-847-195	Sequence 195, App
45	69.4	13.9	593	10	US-09-864-761-7065	Sequence 7065, App

ALIGNMENTS

RESULT 1

US-08-979-847-207

Sequence 207, Application US/08979847

Publication No. US2003039664A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC

APPLICANT: BEDIN, FREDERIC

APPLICANT: PARANHOS-BACCALA, GLAUCIA

APPLICANT: KOMRINOS-PRADTEL, FLORENCE

APPLICANT: JOLIVET-REYNAUD, COLETTE

APPLICANT: MANDRAND, BERNARD

APPLICANT: GARSON, JEREMY

APPLICANT: TUXE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847

FILING DATE: 26-NOV-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: MPB 39046A

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 207:

SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-207

Query Match 86.2%; Score 431.8; DB 7; Length 1600;
Best Local Similarity 90.5%; Pred. No. 5.5e-133;
Matches 459; Conservative 20; Mismatches 22; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTCTATATACGAGCTGTACCTTACCTT 60
DB 373 CAAGAACTCAGGATTATCATGAGGCTGTGTTCTCTATATACGAGCTGTACCTTACCTT 432
QY 61 TATATCTGCTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 433 TATACAGGCTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
QY 121 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 180
DB 493 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 552
QY 181 ACTTCAAAACCAACATCTCACTCACTGATCTTATACCCCAAGGTTCAAGGATAGT 240
DB 553 CTTTGAACCCCAACGTCATCACTGATCTGATCTTATACCCCAAGGTTCAAGGATAGT 612
QY 241 CCCCATCTATTTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACA-- 298
DB 613 CCCCATCTATTTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACA-- 672
QY 299 CTTGCTCTGCTAGTGTGATGATTTATTTGGGCGCCCATTTGAGAACTTGTGCCAT 358
DB 673 CTTGCTCTGCTAGTGTGATGATTTATTTGGGCGCCCATTTGAGAACTTGTGCCAT 732
QY 359 CAAGCAACCCCAAGCGCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 418
DB 723 CAAGCAACCCCAAGCGCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 792
QY 419 AAGGCTCACTGCTGTACAGCAGGT---TACTTAGGGCTAAATTTATCCAAAGGACCC 474
DB 793 AAGGCTCACTGCTGTACAGCAGGT---TACTTAGGGCTAAATTTATCCAAAGGACCC 852
QY 475 AGGGCCCTCAGTGAAGAACATCCAG 501
DB 853 AGGGCCCTCAGTGAAGAACATCCAG 879

RESULT 2

US-08-979-847-206
Sequence 206, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE, PLC
STREET: P. O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA

ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 206:
SEQUENCE CHARACTERISTICS:
LENGTH: 1600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-206

Query Match 85.0%; Score 425.6; DB 7; Length 1600;
Best Local Similarity 88.6%; Pred. No. 6.3e-131;
Matches 449; Conservative 28; Mismatches 24; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTCTATATACGAGCTGTACCTTACCTT 60
DB 373 CAAGAACTCAGGATTATCATGAGGCTGTGTTCTCTATATACGAGCTGTACCTTACCTT 432
QY 61 TATATCTGCTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 433 TATACAGGCTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
QY 121 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 180
DB 493 GATGCTTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 552
QY 181 ACTTCAAAACCAACATCTCACTCACTGATCTTATACCCCAAGGTTCAAGGATAGT 240
DB 553 CTTTGAACCCCAACGTCATCACTGATCTGATCTTATACCCCAAGGTTCAAGGATAGT 612
QY 241 CCCCATCTATTTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACA-- 298
DB 613 CCCCATCTATTTGGGCGAGGATTTAGCCCAAGCTTGAAGCAATCTCTATACCTTGACA-- 672
QY 299 CTTGCTCTGCTAGTGTGATGATTTATTTGGGCGCCCATTTGAGAACTTGTGCCAT 358
DB 673 CTTGCTCTGCTAGTGTGATGATTTATTTGGGCGCCCATTTGAGAACTTGTGCCAT 732
QY 359 CAAGCAACCCCAAGCGCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 418
DB 723 CAAGCAACCCCAAGCGCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 792
QY 419 AAGGCTCACTGCTGTACAGCAGGT---TACTTAGGGCTAAATTTATCCAAAGGACCC 474
DB 793 AAGGCTCACTGCTGTACAGCAGGT---TACTTAGGGCTAAATTTATCCAAAGGACCC 852
QY 475 AGGGCCCTCAGTGAAGAACATCCAG 501
DB 853 AGGGCCCTCAGTGAAGAACATCCAG 879

RESULT 3

US-08-979-847-87
Sequence 87, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-87

Query Match 84.3%; Score 422.2; DB 7; Length 2304;
Best Local Similarity 92.3%; Pred. No. 9.9e-130;
Matches 468; Conservative 0; Mismatches 33; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATATAGCCAGCTGTACCTAGCCCT 60
DB 640 CAAGATCTCAGGATTATTAATGAGGCTGTTTCTCTATATAGCCAGCTGTACCTAGCCCT 699
QY 61 TATATCTGCTTTCCCAATAATCAAGAGAGAGAGGTGTATACAGTCTTGACCTTCAAG 120
DB 700 TATATCTGCTTTCCCTAATAACAGAGAGAGAGGTGTATACAGTCTTGACCTTCAAG 759
QY 121 GATGAGCTCTTCTGATGATCTGATCATCTGATCTCAATCTTGTGGCTTGAAGT 180
DB 760 GATGAGCTTTTCTGATGATCTGATCATCTGATCTCAATCTTGTGGCTTGAAGT 819
QY 181 ACTTGAACCCCAACATCTCAATCTGATCTGATCTATTTTACCCCAAGGTTTGAAGGATACT 240
DB 820 CTTTGAACCCCAACATCTCAATCTGATCTGATCTTATTTTACCCCAAGGTTTGAAGGATACT 879
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTGAGCAATCTCATACCTTGACACA-- 298
DB 880 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTGAGCAATCTCATACCTTGACACT 939
QY 299 CTGTGCTTGGTAGTGATGATTTACTTTGGCCGCCCATTTCAAGAAACCTTGTGCAT 358
DB 940 CTGTGCTTGGTAGTGATGATTTACTTTGGCCGCCCATTTCAAGAAACCTTGTGCAT 999
QY 359 CAAGCACCACCAAGCGCTTCAATTTCTCGGTACCTGTGGGTACATGTTTCCAAACCA 418

DB 1000 CAAGCACCACCAAGCTTCTTACTTCTCTACTACTGCTGCTCAAGGTTTCCAAACCA 1059
QY 419 AAGGCTCAACTCTGCTCAACAGCAG-----GTTACTTAGGGCTTAATAATTCGAAAGCACC 474
DB 1060 AAGGCTGGCTCTGCTCTACAGAGATTAGATTAAGGGCTTAATAATTAATCAAGGACACC 1119
QY 475 AGGGCCCTCAAGTAGGAACACATCCAG 501
DB 1120 AGGGCCCTCAAGTAGGAACGATTCAG 1146

RESULT 4
US-08-979-847-1
Sequence 1, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-1

Query Match 83.8%; Score 419.6; DB 7; Length 1158;
Best Local Similarity 91.9%; Pred. No. 5.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATATCTGCTTTCCCAATAATCAAGAGAGAGAGGTGTATACAGTCTTGACCTTCAAG 120
DB 121 TATATCTGCTTTCCCAATAATCAAGAGAGAGAGGTGTATACAGTCTTGACCTTCAAG 180

QY 121 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATCTTGTTCCTTTGAAGAT 180
DB 181 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATCTTGTTCCTTTGAAGAT 240
QY 181 ACTTCAAAACCAACATCTCAATCTCACTGAGACTATTTTACCCTCAAGGGTTCAAGGATGT 240
DB 241 CCTTTGAACCCCAACCTCTCAATCTCACTGAGACTATTTTACCCTCAAGGGTTCAAGGATGT 300
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCCATCTCTATCTGAGACA-- 298
DB 301 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCATTTCTCATCTGAGACACT 360
QY 299 CTGTCTCTTGGGTAGTGATGATTTACTTTTGTGACCCCATCTCAAGAACTTTGTGCAT 358
DB 361 CTGTCTCTTGGGTAGTGATGATTTACTTTTGTGACCCCATCTCAAGAACTTTGTGCAT 420
QY 359 CAAGCCACCCAGGCGCTCTTCAATTTCTGCGTCTCTGTGGTTACATGTTTCCAAACCA 418
DB 421 CAAGCCACCCAGGAACTTTACTTTCTCACTACCTGTGGTTACAAAGGTTTCCAAACCA 480
QY 419 AAGGCTCACTGTGCTCAGACAG---GTTACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 481 AAGGCTCGGCTGTGCTCAGACAGATTTAGATCTTAGGGCTAAATTTATCCAAAGGACAC 540
QY 475 AAGGCTCTCAGTAGAGAAACATCCAG 501
DB 541 AAGGCTCTCAGTAGAGAAACATCCAG 567

RESULT 5

US-08-979-847-205

; Sequence 205, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
APPLICANT: BESENE, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPA 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 205:
SEQUENCE CHARACTERISTICS:
LENGTH: 1597 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-205

Query Match 83.8%; Score 419.6; DB 7; Length 1597;
Best Local Similarity 91.9%; Pred. No. 6.1e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCCAGAGATTATCAATAGAGCTGTGTCTCTATATGACGAGCTGTACAGCCCT 60
DB 373 CAAGATCCAGAGATTATCAATAGAGCTGTGTCTCTATATGACGAGCTGTACAGCCCT 432
QY 61 TATATCTGCTTTCCCAATATCCAGAGAAACAGAGTGTATTACAGTCTGACCTTCAG 120
DB 433 TATATAGTCTTTCCCAATATCCAGAGAAACAGAGTGTATTACAGTCTGACCTTCAG 492
QY 121 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATCTTGTGTGCTTTGAAGAT 180
DB 493 GATGCTTCTTCTGATCCCTGTACATCTGATCTCTCAATCTTGTGTGCTTTGAAGAT 552
QY 181 ACTTCAAAACCAACATCTCAATCTCACTGAGACTATTTTACCCTCAAGGTTTCAAGGATGT 240
DB 553 CTTTGAACCCCAACGTCTCAATCTCACTGAGACTATTTTACCCTCAAGGTTTCAAGGATGT 612
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCCATCTCATACCTGAGACACT 298
DB 613 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCCATCTCATACCTGAGACACT 672
QY 299 CTGTCTCTTGGGTAGTGATGATTTACTTTTGGCCCATCTCAAGAACTTTGTGCCAT 358
DB 673 CTGTCTCTTGGGTAGTGATGATTTACTTTTGGCCCATCTCAAGAACTTTGTGCCAT 732
QY 359 CAAGCCACCCAGGCGCTTTCAATTTCTGCGTCACTGTGCTCAATGTTTCCAAACCA 418
DB 733 CAAGCCACCCAGGAACTTTACTTTCTCACTGTGCTCAATGTTTCCAAACCA 792
QY 419 AAGGCTCACTGTGCTCAGACAG---GTTACTTAGGGCTAAATTTATCCAAAGGACAC 474
DB 793 AAGGCTCGGCTGTGCTCAGACAGATTTAGATCTTAGGGCTAAATTTATCCAAAGGACAC 852
QY 475 AAGGCTCTCAGTAGAGAAACATCCAG 501
DB 853 AAGGCTCTCAGTAGAGAAACATCCAG 879

RESULT 6

US-08-979-847-88

; Sequence 88, Application US/08979847

; Publication No. US20030039664A1

; GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
APPLICANT: BESENE, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-2787
TELEFAX: 703-836-6400
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 2365 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-88

Query Match 83.8%; Score 419.6; DB 7; Length 2365;
Best Local Similarity 91.9%; Pred. No. 7.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGTATTAATGAGGCTGTGTTCTCTATAGCCGCTGTACCTACCT 60
DB 646 CAAGAACTCAGAGTATTAATGAGGCTGTGTTCTCTATAGCCGCTGTACCTACCT 705
QY 61 TATACCTGCTTTCCCAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACAGGCTTTCCCAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTCTTCTGTCATCCCTGTACATCTGATCTCAATCTTGTGCTTTGAAGAT 180
DB 766 GATGCTCTTCTGTCATCCCTGTACATCTGATCTCAATCTTGTGCTTTGAAGAT 825
QY 181 ACTTCAAAACCAACATCTCACTGACATATTTTACCCTCAAGGATGAT 240
DB 826 CTTTGAACCCCAAGCTCACTGACATATTTTACCCTCAAGGATGAT 885
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCTTCACTACCTGACCA-- 298
DB 886 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCTTCACTACCTGACCACT 945
QY 299 CTTGCTCTTGGTAGTGATGATTTACTTTTGGCCGCCATTGCAAACTTGTGCCAT 358
DB 946 CTTGCTCTTGGTAGTGATGATTTACTTTTGGCCGCCATTGCAAACTTGTGCCAT 1005
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTGATAGTGTTCAAACCA 418
DB 1006 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTGATAGTGTTCAAACCA 1065
QY 419 AAGGCTCAACTCTGCTCAGAGAG--GTTACTAGAGGCTAAATTTTCAAAAGGACCC 474
DB 1066 AAGGCTCAACTCTGCTCAGAGAGAGTTAGATTTAGAGGCTAAATTTTCAAAAGGACCC 1125
QY 475 AAGGCTCTCAGTGAAGAAACATCCAG 501
DB 1126 AAGGCTCTCAGTGAAGAAACATCCAG 1152

RESULT 7
US-08-979-847-53
Sequence 53, Application US/08979847
Publication No. US200300396441
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESENE, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA

APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GABSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-2787
TELEFAX: 703-836-6400
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-53

Query Match 83.8%; Score 419.6; DB 7; Length 2391;
Best Local Similarity 91.9%; Pred. No. 7.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGTATTAATGAGGCTGTGTTCTCTATAGCCGCTGTACCTACCT 60
DB 646 CAAGAACTCAGAGTATTAATGAGGCTGTGTTCTCTATAGCCGCTGTACCTACCT 705
QY 61 TATACCTGCTTTCCCAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACAGGCTTTCCCAATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTCTTCTGTCATCCCTGTACATCTGATCTCAATCTTGTGCTTTGAAGAT 180
DB 766 GATGCTCTTCTGTCATCCCTGTACATCTGATCTCAATCTTGTGCTTTGAAGAT 825
QY 181 ACTTCAAAACCAACATCTCACTGACATATTTTACCCTCAAGGATGAT 240
DB 826 CTTTGAACCCCAAGCTCACTGACATATTTTACCCTCAAGGATGAT 885
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCTTCACTACCTGACCA-- 298
DB 886 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCTTCACTACCTGACCACT 945
QY 299 CTTGCTCTTGGTAGTGATGATTTACTTTTGGCCGCCATTGCAAACTTGTGCCAT 358
DB 946 CTTGCTCTTGGTAGTGATGATTTACTTTTGGCCGCCATTGCAAACTTGTGCCAT 1005
QY 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTGATAGTGTTCAAACCA 418
DB 1006 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTGATAGTGTTCAAACCA 1065

PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 16537
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE: MAP TO AC007567.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
US-09-864-761-16537

Query Match 80.3%; Score 402.2; DB 10; Length 568;
Best Local Similarity 91.1%; Pred. No. 2.3e-122; Indels 7; Gaps 3;
Matches 462; Conservative 0; Mismatches 38;

QY 1 CAAGATCTCAGGATTATCATGAGGCTGTTCTCTTATATACGAGCTGTACCTAGCCCT 60
DB 58 CAAGATCTCAGGATTATCATGAGGCTGTTCTCTTATATACGAGCTGTACCTAGCCCT 117
QY 61 TATACTGCTCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 118 TATACTGCTCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
QY 121 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 180
DB 178 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGATCTG 237
QY 181 ACTTGAACCCCAATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTG 240
DB 238 CTTTGAACCCCAATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTG 297
QY 241 CCCCATTATTTGGCCAGGATTTAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 298
DB 298 CCCCATTATTTGGCCAGGATTTAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 357
QY 299 CTTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 358 CTTTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
QY 359 CAAGCACCAGGAGCTCTTCAATTTCTGCTACCTGAGGCTACATGAGTTCCAAACCA 418
DB 418 CAAGCACCAGGAGCTCTTCAATTTCTGCTACCTGAGGCTACATGAGTTCCAAACCA 477
QY 419 AAGGCTCACTGCTGCTACAGCAGG---TTACTTAGGCTTAAATTTATCCAAAGGAC 474
DB 478 AA-GCTCAACTGCTGCTACAGCAGGCTTAAATTTATCCAAAGGAC 536
QY 475 AGGGCCCTCAGTAGGAGACATCCAG 501
DB 537 AGGGCCCTCAGTAGGAGACATCCAG 563

RESULT 10
US-09-864-761-11981
Sequence 11981, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11981
LENGTH: 557
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007923.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-11981

Query Match 47.2%; Score 236.6; DB 10; Length 557;
Best Local Similarity 72.2%; Pred. No. 2e-68;
Matches 337; Conservative 0; Mismatches 124; Indels 6; Gaps 2;

QY 41 AGCCAGCTGATCTAGCCCTTATATCTGCTTTCCCAATATCCAGAGAGAGAGAGAG 100
DB 1 ACCCAGTTATACCCAAACCCCTGATCCCTGCTCTCAAAATACAGAGAGAGAGAGAT 60
QY 101 TTACAGTCTGAGACCTTAGAGATGCTTTCTTGATCCCTGTACATCCCTACTCAAT 160
DB 61 TCACGTCTTAGACCTCAAGATGCTTTCTGTATCCCTGACTTACCTCCAGT 120
QY 161 TCTTGTGCTTTGAAGATATCTTCAACCAACCAATCTCACTCACTGACTATTTTAC 220
DB 121 TTCTCTTCTGCTTTGAGATCCACAGACACACATCCCACTTATGTGACAGCTTGC 180
QY 221 CCAAGGCTCAGAGATATGCTCCCATCTATTGTCAGGCACTTACCCCAAGCTTGACC 280
DB 181 CCAAGGCTTATAGGATAGCCCTCATCTGTTGGTCAGGCACTGAGCCCAAGATCTAGACC 240
QY 281 AATCTCA--TACCTGACACTTGTCTGCTGCTGAGTGAATTTACTTTGGCCGCC 338
DB 241 ACTTCTAAGTCCAGGACCTGCTGCTTCTTCAATGATGATGATTTACTTTGGCTACCA 300

Qy 339 ATTCAAAACCTTGCCATCAAGCCACCAGCGCTTCAATTCTCTGCTACTGTG 398
Db 301 GTTTGGAAGCCCTCATGTGAGCAGGCTACTCTGATCTTGAATCTTCTAGTATCAAG 360
Qy 399 GGTACATGTTTCCAAACCAAGGCTCAACTGTG-TTCACAGAGGT---TACTTAGGGC 454
Db 361 GGTACAGGCACTGTAGTGTGAAGGCTCAGCTTTCCTACAGCAGGCTCAAAATCTTAGGCGC 420
Qy 455 TAAATTTATCCAAAGCAGCAGGCGCTCTGATGTAACACATCCAG 501
Db 421 TATCTTATCCAGAGGAGCAGGCGCTCTGACGACGAATGATACAG 467
RESULT 11
US-09-864-761-2921/c
Sequence 2921, Application us/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonex Sequence Listing Engine ver. 1.1
SEQ ID NO 2921
LENGTH: 449
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010856.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
US-09-864-761-2921
Query Match 46.2%; Score 231.6; DB 10; Length 449;
Best Local Similarity 72.8%; Pred. No. 8.4e-67;
Matches 326; Conservative 0; Mismatches 119; Indels 3; Gaps 2;
Qy 41 AGCCAGCTGTACTTACGCTTATCTGCTTTCCAAATACAGAGAGACAGAGTGT 100
Db 449 ATCCAGCTATACCAACCCCTATACCGCTCTCTCAAAATACAGAGAGACAGAGTGT 390
Qy 101 TTACAGTCTGACCTTCAGAGATGCTTTCTTGATCCCTGTACATCTGACTCTCAAT 160
Db 389 TCATTTGTTGACCTCAAGAGATGCTTCCCTGCTCATTTCTCTGACCTGACCTCAAG 330
Qy 161 TCTGTTGCTTTGAAGATCTTCAACCAACATCTCAACCTGACCTGACTTTTAC 220
Db 329 TTCTTTGCTTTGAGATCCACAGACACAGCTCCCACTTCTGATGCTTTGCT 270
Qy 221 CCAAGGCTTCAGAGATAGTCCCATCTATTGGCCAGGCACTTACCCCAAGCTTGAGCC 280
Db 269 CTCAAGGCTTAGGAGATAGCTTATCTGTTGTGACAGGACTGCGCCCAAGATCTAGGCC 210
Qy 281 AATCTCA--TACTGCACTTGTCTTGGTGGTGGATGATTTACTTTTGGCCGCC 338
Db 209 ACTTCTCAAGTACAGGACCTGCTCTGCTGATGAGATTTACTTATGTTACCA 150
Qy 339 ATTCAAAACCTTGCCATCAAGCCACCAGCGCTTCAATTCTCTGCTACTGTG 398
Db 149 GTCAAAAGCTCATGCGCAGCAGGCTTCTAGCTCTTCTTAACCTTCTAGCTATACAG 90
Qy 399 GGTACATGTTTCCAAACCAAGGCTCAACTGTGCTGAC-AGCAGTTACTTAGGGCTAA 457
Db 89 GGTACATGCTATTAATGTAAGGCCAGGCTCTGCTGACGAATATATAGGCGCTAA 30
Qy 458 AATATCCAAAGCAGCAGGCGCTCTAG 485
Db 29 TCTTAGCAGAGGAACGAGGCTCTAG 2
RESULT 12
US-09-864-761-1194/c
Sequence 1194, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30


```

; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1194
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006999.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.3
; OTHER INFORMATION: EXPRESSED IN B1474, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
; US-09-864-761-1194

Query Match 44.2%; Score 221.6; DB 10; Length 467;
Best Local Similarity 70.7%; Pred. No. 1.8e-63;
Matches 325; Conservative 0; Mismatches 129; Indels 6; Gaps 2;

```

```

Db 107 GCTGTTAGTCAAGGCCAGCTTGTGCTACAGCAGGCCAATATCTAGCTTAATCAT 48
Qy 462 ATCAAGGACGACGAGGCCCTCAGTGAAGAACATCCAG 501
Db 47 AGCCAGAGGACCAAGGCCCTCAGCAGAGGAATGAAAAAG 8

RESULT 13
US-08-979-847-2
; Sequence 2, Application US/08979847
; Publication No. US2003039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TURE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-2

Query Match 42.4%; Score 212.6; DB 7; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.4e-60;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```



```

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 645 base pairs
;   TYPE: nucleotide
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
; US-08-979-847-8

```

```

Query Match      42.2%; Score 211.4; DB 7; Length 645;
Best Local Similarity 89.9%; Pred. No. 5e-60;
Matches 250; Conservative 0; Mismatches 22; Indels 6; Gaps 2;

```

```

QY 230 TCAGGATAGTCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 289
Db 1 TCAGGATAGCCCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 60
QY 290 ACCTGGACA--CTGTCTCTGGGTAAGTGATGATTACTTTGGCCGCCCATTCAGAAA 347
Db 61 ACCTGGACACTCTGTCTCTCTGATGATGATGATTACTTTAGTCCGCCGTTTCAGAAA 120
QY 348 CCTGTGCGCATCAAGGCCACCAAGGCTCTCAATTTCTGCTACTGTTGGCTACATGG 407
Db 121 CCTGTGCGCATCAAGGCCACCAAGGCTCTCAATTTCTGCTACTGTTGGCTACATGG 180
QY 408 TTTCCAAACCAAAGGCTCACTGTGCTCAGCAG---GTTACTTAGGGCTAAATTAT 463
Db 181 TTTCCAAACCAAAGGCTGTGGCTGTCTCAGAGGATTAGATACCTAGGCTAAATTAT 240
QY 464 CCAAGGACCAAGGGGCTCTCAGTGAGGAACACATCCAG 501
Db 241 CCAAGGACCAAGGGGCTCTCAGTGAGGAACGATCCAG 278

```

Search completed: April 17, 2003, 07:46:44
 Job time : 127.415 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:30:25 ; Search time 1268.84 seconds
(without alignments)
6394.748 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 1 caagatccagcagatcatca.....tcagtgaggaacacatccag 501

Sequence: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Scoring table: 16154066 seqs, 8097743376 residues

Searched: Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	384.2	76.7	521	17	AZ517418 RPCI-11-4
2	349	69.7	674	17	A0111469 CIT-HSP-2
3	324.2	64.7	533	17	B17809 34719.TVB
4	300	59.9	594	17	A0193544 CIT-HSP-2
5	266.8	53.3	595	17	B59017 CIT-HSP-201
6	265.6	53.0	745	9	AU121943 AU121943

C 7	256.6	51.2	750	9	AL698185
C 8	253.2	50.5	231	12	BF989696
C 9	252.6	50.4	707	17	B66771
C 10	208	41.5	736	10	AV731083
C 11	206.2	41.2	537	17	A0270891
C 12	204.4	40.8	416	17	A0041889
C 13	204.4	40.8	518	17	AQ361037
C 14	203.8	40.7	434	10	BE168410
C 15	201	40.1	408	17	AQ033042
C 16	190	37.9	549	17	AQ231407
C 17	177.8	35.5	405	17	AQ033015
C 18	177.2	35.4	557	17	AQ061550
C 19	175.8	35.1	676	9	AL705035
C 20	168	33.5	518	17	AQ243044
C 21	164.8	32.9	428	17	AQ132081
C 22	161.6	32.3	442	17	AQ075734
C 23	159.2	31.8	541	17	AQ0702212
C 24	157.2	31.4	431	17	B17135
C 25	155	30.9	431	17	AQ021075
C 26	152.4	30.4	442	17	AQ0701392
C 27	150.6	30.1	467	17	AQ0876467
C 28	148	29.5	456	17	AQ232635
C 29	148	29.5	539	17	AQ153050
C 30	147.8	29.5	449	17	AQ0701255
C 31	146	29.1	526	17	AQ217989
C 32	143	28.5	332	17	AQ024733
C 33	140.6	28.1	737	17	B67137
C 34	140.4	28.0	437	17	B47870
C 35	140.2	28.0	461	17	AQ225099
C 36	138.4	27.6	518	17	AQ0807540
C 37	132.6	26.5	518	17	AQ284525
C 38	132.2	26.4	429	17	AQ028930
C 39	131.4	26.2	431	17	AQ0736815
C 40	127.2	25.4	433	17	AQ145803
C 41	123	24.6	469	17	AQ223144
C 42	121	24.2	338	17	AQ044697
C 43	119.8	23.9	533	17	AQ226226
C 44	119.8	23.9	533	17	AQ479087
C 45	118.2	23.6	383	17	AQ192721

ALIGNMENTS

RESULT 1
AZ517418/c 521 bp DNA linear GSS 16-OCT-2000
LOCUS RPCI-11-4708-TV RPCI-11 Homo sapiens genomic clone RPCI-11-4708,
DEFINITION
DNA sequence.
ACCESSION AZ517418
VERSION AZ517418.1 GI:10826891
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 521)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.V.C.
TITLE BAC end sequences of library RPCI-11
JOURNAL Unpublished (1997)
COMMENT Other GSSs: RPCI11-4708-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com) . BAC end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
This BAC end was generated during the Rad process and may have
higher chance of clone tracking errors.

Class: BAC ends.
Seq primer: 17

FEATURES
source
location/Qualifiers

1..521
/organism="Homo sapiens"
/db_xref="GDB:7518007"
/db_xref="taxon:9606"
/clone="RPC1-11-4708"
/clone_11b="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

BASE COUNT 145 a 93 c 147 g 136 t
ORIGIN

Query Match 76.7%; Score 384.2; DB 17; Length 521;

Best Local Similarity 90.8%; Pred. No. 1e-111; Mismatches 38; Indels 6; Gaps 2;

Matches 433; Conservative 0; Mismatches 38; Indels 6; Gaps 2;

Oy 31 GTTCCCTATAGCCAGCTGTACCTAGCCCTTACTGTGCTTCCCAATACAGAGNA 90
Db 521 GTTCCCTATAGCCAGCTGTACCTAGCCCTTACTGTGCTTCCCAATACAGAGNA 462
Oy 91 GCAGAGTGTATACAGTCTGACCTTCAGATGCTTCTTCTGCATCCCTGTACATCT 150
Db 461 GCAGAGTGTATACAGTCTGACCTTCAGATGCTTCTTCTGCATCCCTGTACATCT 402
Oy 151 GACCTGCATCTGTGTGCTTGGAGATGATCTCAACCAACATCTCACTGACCTGG 210
Db 401 GACCTGCATCTGTGTGCTTGGAGATGATCTCAACCAACATCTCACTGACCTGG 342
Oy 211 ACTATTTTACCCCAAGGCTTCAGGATAGTCCCATC..ATTGGCCAGGCAATTAGCCCA 270
Db 341 ACTGTTTACCCCAAGGCTTCAGGATAGTCCCATC..ATTGGCCAGGCAATTAGCCCA 282
Oy 271 GACTTGAGCAATCTCTCATACCTGACA--CTTGTCTTGGGTAGGTGATATTACTT 328
Db 281 GACTTGAGCAATCTCTCATACCTGACAATCTTGTCTTCAATAGGGATATTAAAT 222
Oy 339 TTGGCGGCCCATTCAGAAACCTTGGCCATCAAGCACCAAGCGCTTCAATTTCTC 388
Db 221 TTGGCGGCCCATTCAGAAACCTTGGCCATCAAGCACCAAGCGCTTCAATTTCTC 162
Oy 389 GCTACCTGTGCTAGATGTTTCCAAACCAAGGCTCAACTGTCTCAACGAGGT--- 444
Db 161 GCTACCTGTGCTAGATGTTTCCAAATGAAAGGCTCA..CTGTGCTCAACGAGGTAA 102
Oy 445 TACTTAGGGCTAAATTTATCCAAAGGACACAGGCGCTT..AGTGAGGACACATCCAG 501
Db 101 TACTTAGGGCTAAATTTATCCAAAGGACACAGGCGCTT..AGTGAGGAAATCATCCAG 45

RESULT 2
AQ111469/c 674 bp DNA linear GSS 29-AUG-1998
LOCUS CIT-HSP-2372020.TR CIT-HSP Homo sapiens genomic clone 2372020, DNA
DEFINITION
ACCESSION AQ111469
VERSION AQ111469.1 GI:3488126
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 674)
Adams,M.D., Rounley,S.D., Zhao,S., Bass,S., Linhar,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready

JOURNAL Map Building
Unpublished (1998)

COMMENT Other GSSs: CIT-HSP-2372020.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: madams@tigr.org

Clones are available from Research Genetics (info@resgen.com) . BAC

end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

FEATURES
source
location/Qualifiers

1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2372020"
/clone_11b="CIT-HSP"
/sex="Male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 180 a 128 c 184 g 182 t
ORIGIN

Query Match 69.7%; Score 349; DB 17; Length 674;

Best Local Similarity 90.3%; Pred. No. 2.1e-100; Mismatches 411; Conservative 0; Mismatches 30; Indels 14; Gaps 3;

Oy 53 CTAGCCCTTATACCTGCTTCCCAATATACAGAGGAGAGGCTTACAGTCTGG 112
Db 674 CTAGCCCTTATACCTGCTTCCCAATATACAGAGGAGAGGCTTACAGTCTGG 615
Oy 113 ACCTTACAGATGCTTCTTCTGCATCTCCCTGTACATCTGATCTCAATCTTGTGCT 172
Db 614 ACCTTACAGATGCTTCTTCTGCATCTCCCTGTACATCTGATCTCAATCTTGTGCT 555
Oy 173 TTGAAGATCTTCAAAACCAATCTCACTCACTGACCTGACATTTTACCCCAAGGTTCA 232
Db 554 TTGAAGATCTTCAAAACCAATCTCACTCACTGACCTGACATTTTACCCCAAGGTTCA 495
Oy 223 GGGATAGTCCCATCTATTTGGCCAGGATTAAGCCAAAGCTTGGAGCAATCTCATAC 292
Db 494 GGGATAGTCCCATCTATTTGGCCAGGATTAAGCCAAAGCTTGGAGCAATCTCATAC 435
Oy 293 TGACACA--CTTGTCTTGGGTAGGTGATGATTTACTTTGGCCGCTTACAGAACT 350
Db 434 AGGACACTCTTGTCTTCAAGTATGATGATTTACTTTTGGCCGCTTACAGAACT 375
Oy 351 TGTGCGATCAAGCCACCAAGCGCTTCAATTTCTGCTGCTACCTGTGCTACATGTTT 410
Db 374 TGTGCGATCAAGCCACCAAGCGCTTCAATTTCTGCTGCTACCTGTGCTACATGTTT 323
Oy 411 CCAAGCAAGGCTCAACTGTCTCAACGAGGT---TACTTAGGGCTAAATATATCA 466
Db 322 CCAAGCAAGGCTCAACTGTCTCAACGAGGTAAATATATATATATATATATCA 263
Oy 467 AAGGACACAGGCGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 501
Db 262 AAGGACACAGGCGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 228

RESULT 3
B17809/c 533 bp DNA linear GSS 04-JUN-1998
LOCUS B17809
DEFINITION 347L9.TVB CIT978SKA1 Homo sapiens genomic clone A-347L9, DNA
ACCESSION B17809
VERSION B17809.1 GI:2125558

KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 533)
AUTHORS Adams,M.D., Kelley,J.M., Rounsley,S.R. and Venter,J.C.
TITLE Use of a BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL Unpublished (1997)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

FEATURES
source
1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-347L09"
/clone_lib="CIT978SKA1"
/sex="Female"
/cell_type="Fibroblast"
/note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A1"
BASE COUNT 154 a 97 c 144 g 138 t
ORIGIN

Query Match 64.7%; Score 324.2; DB 17; Length 533;
Best Local Similarity 86.0%; Pred. No. 1.6e-92;
Matches 398; Conservative 0; Mismatches 53; Indels 12; Gaps 3;

QY 50 TACCCAGCCCTTATCTCGCTTCCCAATACGAGGAGCAAGTGTATACAGTCC 109
DB 533 TACTTAACCCCTTATCTCGCTTCCCAATACGAGGAGCAAGTGTATAGTCT 474
QY 110 TGGACCTTCAGAGTCCCTTCTTCGATCCCTGTACATCCCTCAATCTGTTTG 169
DB 473 TGGACCTTAAAGATGCTTTTCTGATTCCTGTACATCCCTCAATCTGTTTG 414
QY 170 CCTTGAAGATACCTTCAACCCCAATCTCACTGACCTGACATTTTACCCCAAGGT 229
DB 413 CCTTGAAGATCCTTCAACCCCAATCTCACTGACCTGACATTTTACCCCAAGGT 354
QY 230 TCAGGATATGTCCTATTTGGCCAGGAT-----TAGCCCAAGCTTGAAGCAT 283
DB 353 TCAGGATATGTCCTATTTGGCCAGGATCAACCCCAAGCTTGAAGCAT 294
QY 284 CCTCACTCGGACACT--TGCTTCGGTAGAGGATATTTACTTTGGCCGCAAT 341
DB 293 TCTCATATCTGAACACTATGTTGTTGTTGATATGATATTTAATCTAGCCGCTG 234
QY 342 CAGAAACCTTGTGCATCAAGCCCAAGCGCTCTTCAATTTCTCGTACCTGTGCT 401
DB 233 CAGAAACCTTGTGCATCAAGCCCAAGCGCTTTTAACTTCTCACCACTGTGGCT 174
QY 402 ACATGTTTCCAAACCAAGGCTCACTGTCTCAAGAGT---TACTTAGGCTTAA 457
DB 173 ACATGTTTCCAAACCAAGGCTCACTGTCTTAAAGAGTAAATCTTAGGCTTAA 114
QY 458 AATTATCCAAAGCAACGAGGCTCAGTAGAGAAACATCA 500
DB 113 AATTATCCAAAGCAACGAGGCTCAGTAGAGATGTATCA 71

RESULT 4
A0193544/c

LOCUS A0193544 594 bp DNA linear GSS 16-SEP-1998
DEFINITION CIT-HSP-2381P7.TR CIT-HSP Homo sapiens genomic clone 2381P7, DNA
sequence.
ACCESSION A0193544
VERSION A0193544.1 GI:3603659
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Baes,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL Unpublished (1998)
COMMENT Other GSSs: CIT-HSP-2381P7.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source
1..594
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2381P7"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 151 a 111 c 170 g 162 t
ORIGIN

Query Match 59.9%; Score 300; DB 17; Length 594;
Best Local Similarity 91.7%; Pred. No. 9.4e-85;
Matches 341; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 136 ATCCCTGTACATCTGACTCTCAATCTTGTGCTTGAAGATCTTCAACCAACA 195
DB 594 ATCCCTGTACATCTGACTCTCAATCTTGTGCTTGAAGATCTTCAACCAACA 535
QY 196 TCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGTCCCATCTATTGGC 255
DB 534 TCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGTCCCATCTATTGGC 475
QY 256 CAGGATATGCCCAAGACTTGAAGCAATCTCATCTGACA--CTGTCTCTTGCTAG 313
DB 474 CAGGATATGCCCAAGACTTGAAGCAATCTCATCTGAGACCTTGTCTTGCTAG 415
QY 314 GTGATGATTTACTTTTGGCCGCGCATTCAGAAACCTTGTGCATCAAGCCCAAGCG 373
DB 414 ATGATGATTTATTTTATTTAGCTGCGCGGTCAAAAACCTTGTGCATCAAGCCCAAGCG 355
QY 374 CTCTTCAATTTCTCTGCTACTGTGAGCTACATGTTTCCAAACCAAGGCTCAATCTGCG 433
DB 354 CTCTTCAATTTCTCTGCTACTGTGAGCTACATGTTTCCAAACCAAGGCTCAATCTGCG 295
QY 434 TCACAGCAGGT---TACTTAGGGCTAAATATTCACCAAGGCAACGAGGCTCAGTAG 489
DB 294 TCACAGCAGGTTAAATTTAGGGCTAAATATTCACCAAGGCAACGAGGCTCAGTAG 235
QY 490 GAACACATCCAG 501

Db 234 GAACGTTCCAG 223

RESULT 5
B59017/c 595 bp DNA linear GSS 20-JUN-1998
LOCUS CIT-HSP-2014K19.TR CIT-HSP Homo sapiens genomic clone 2014K19, DNA
DEFINITION sequence.
ACCESSION B59017
VERSION B59017.1 GI:2613735
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 595)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Base,S., Linher,K., Golden
,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: CIT-HSP-2014K19.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Classes: BAC ends.

FEATURES
source 1..595
location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7042821"
/db_xref="taxon:9606"
/clone="2014K19"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

BASE COUNT 159 a 104 c 159 g 174 t
ORIGIN

Query Match 53.3%; Score 266.8; DB 17; Length 595;
Best Local Similarity 89.2%; Pred. No. 3.9e-74;
Matches 323; Conservative 0; Mismatches 32; Indels 7; Gaps 3;

QY 146 ATCTGACTCTCAATTCCTGTTGCTTGAAGACTTCGAACCAACCAACATCTCACTCA 205
Db 595 ATCCGAGCTCTCAATTCCTGTTGCTTGAAGACTTCGAACCAACCAACATCTCACTCA 536
QY 206 CCGTGAATTTTACCCCAAGGATTCAAGGATAGTCCCATATTTTGGCCAGGATTAG 265
Db 535 CCGTGAATTTTACCCCAAGGATTCAAGGATAGTCCCATATTTTGGCCAGGATTAG 476
QY 266 CCGAAGACTTGAACCAATCTCATACCTGACACT-TTCTCTGGTAGTGATGATT 323
Db 475 CTCAAGACTTGAACCAATCTCATACCTGACACTTCTTTGAGACATGATGATT 416
QY 324 TACTTTGGCCGCAATTCAGAACTTGTGCATCAAC CCAACCAAGCGCTTCAATT 383
Db 415 TACTTTAGTCCCTTTGAGAACTTGTGCATCAAC CCAACCAAGCATTTTAATT 356
QY 384 TCCTGCTTACCTGTGGCTACATGTTTCCAAACCAAGGCTCAACTGCTCAACGACG 443
Db 355 TCCTCTCACCTGTG-GATTAAGGTTTCCAAACCAAGGATGAGCTGCTCAACGATAG 297

QY 444 T----TACTTAGGGCTAAATTAATTCAAAGGACACAGGCGCTCAGTAGGAACATCC 499
Db 296 TTAATACTTAGTCTAAATTAATTCAAAGGACACAGGCGCTCAGTAGGAACATATCC 237

QY 500 AG 501
Db 236 AG 235

RESULT 6
AUI21943 745 bp mRNA linear EST 01-AUG-2002
LOCUS AUI21943 MAMMA1 Homo sapiens cDNA clone MAMMA1001324 5', mRNA
DEFINITION sequence.
ACCESSION AUI21943
VERSION AUI21943.1 GI:10937178
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
location/Qualifiers
1..745
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMA1001324"
/clone_1lb="MAMMA1"
/tissue_type="mammary gland"
/note="Vector: pME18SFL3"

BASE COUNT 214 a 196 c 156 g 176 t 3 others
ORIGIN

Query Match 53.0%; Score 265.6; DB 9; Length 745;
Best Local Similarity 89.8%; Pred. No. 1.1e-73;
Matches 309; Conservative 0; Mismatches 29; Indels 6; Gaps 2;

QY 164 TGTTCGCTTTGAAGACTTCAACCAACCAACATCTCAACTCACTGAGACTATTTTACCC 223
Db 1 TGTTCGCTTTGAAGACTTCAACCAACCAACATCTCAACTCACTGAGACTATTTTACCC 60
QY 224 AAGGTTCAAGGATAGTCCCATCAATTTGGCCAGGCTTAGCCCAAGCTTGAACCAAT 283
Db 61 AAGGTTCAAGGATAGTCCCATCAATTTGGCCAGGCTTAGCCCAAGCTTGAACCAAT 120
QY 284 CCTATACCTGAGCA--CTTGTCTTGGTAGTGATGATTTCCTTTGGCCGCCATT 341
Db 121 TCTCACACTGAGCACTTGTCTTGTGATGATGATGATTTACTTACCTGCCCTT 180
QY 342 CAGAAACCTTGTGCATCAACCAACCAAGCGCTTCAATTTCTGCTACCTGTGCT 401
Db 181 CAGAAACCTTGTGCATCAACCAACCAAGCGCTTCAATTTCTGCTACCTGTGCT 240
QY 402 ACATGCTTTCAAACCAAGGCTCACTGCTCAACGACG----TTACTTAGGGCTAA 457
Db 241 ACAGGTTTCAAACCAACCAAGGCTCACTGCTCAACGACGCTTAATTAATTAAGGCTAA 300

QY 458 AATTATCCAAAGCAGCAGGCGCTCAGTGAGGAACATCATCAG 501
Db 301 AATTATCTTAGGACAGGCGCTCAGTGAGGAATGATCCAG 344

RESULT 7
AL698185/c 750 bp mRNA linear EST 21-MAR-2002
LOCUS DKEZP686N20106.r1.686 (synonym: hlc3) Homo sapiens cDNA clone
DEFINITION DKEZP686N20106.5', mRNA sequence.
ACCESSION AL698185
VERSION AL698185.1 GI:19618725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS Ottenwaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Weill, B. and Wiemann, S.
TITLE EST (Ottenwaelder, B., Obermaier, B., Mewes, H.W., Weill, B. and Wiemann, S.)
JOURNAL Unpublished (2001)
COMMENT Contact: Ottenwaelder B
MIPS
Am Klopferstritz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
Sequenced by Medigenomix (Martinsried/Germany) within the CDNA
sequencing consortium of the German Genome Project. No si sequence
available.
This clone (DKEZP686N20106) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg; GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKEZP686N20106"
/clone_lib="686 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex; Site_1: SfiI; Site_2: SfiI; B;
CDNA-collection"

BASE COUNT 195 a 151 c 194 g 210 t

ORIGIN
Query Match 51.2%; Score 256.6; DB 9; Length 750;
Best Local Similarity 73.1%; Pred. No. 8.2e-71;
Matches 372; Conservative 0; Mismatches 129; Indels 8; Gaps 3;

QY 1 CAAGATCTCAGATTTATGATGAGGCTGTCTCTCTATAGCCAGTGAAGT--AGCC 58
Db 582 CAAGATCTTACATATGATGAGGAGTAAATCTCTATATGAGCTGTACCCCAACC 523

QY 59 CTTATCTCTGCTTTCCCAATACCAAGAGAGAGAGAGTGGTTTACATCTCGACCTTC 118
Db 522 CTTATACCTCTCTCTCAATATACCAAGAGAGAGAGTGGTTTCTTTGACACCTCA 463

QY 119 AGGATGCTCTCTCTGATCCCTGATACCTGATCTCAATCTCTGTTGGCTTTGAG 178
Db 462 AGGATGCTCTCTCTGATCCCTGATACCTGATCTCTGCTCTCTTTGGCTTTGAG 403

QY 179 ATACTTCAACCCCAATCTCACTCACTGATGATATTTTACCCCAAGGGTTCAAGGATA 238
Db 402 ATCCACAGACACACGTCCTCAATCATGAGCGGTCTTGCTCAAGGTTTAGGATA 343

QY 239 GTCCCATCTATTTGGCAGGATTTAGCCCAAGATTTAGCCCAATCTCA--TACTTGA 296
Db 342 GCCCTAATCTGTTGGTCAAGGACCTGGCCCAAGATCTAGGCCCACTTCAAGTTCAGGACA 283

QY 297 CACTGTCTCTGCTAGGAGTGAATGATTTACTTTGGCGGCCCATTCAGAAACCTTGTGCC 356
Db 282 CTGTGTCCTTCAAGTATGATGATTTACTTTGGCTACAGGTTCAAGAGCTTCATGCC 223

QY 357 ATCAAGCACCACAGCCCTCTTCAATTTCTCGCTACCTGTGGCTACATGTTTCCAAAC 416
Db 222 AGCAGGCTACTCTGATCATCTGAACTTTCTAGCTAATCAAGGGGTACAGGCGCTTAAT 163

QY 417 CAAGGCTCAACTCTGCTCAACAGAGGT---TACTTAGGGCTAAATATTCAGAGCA 472
Db 162 TGAAGGCCAGCTCTGCTCTACACAGATCAATATCTTAGGCTTAATCTTAGCCAGAGGA 103

QY 473 CCAGGCGCTCTGATGAGGAACATCCAG 501
Db 102 CCAACACCTCAGCAAGAAATGAAGACAG 74

RESULT 8
BF989696/c 291 bp mRNA linear EST 23-JAN-2001
LOCUS BF989696
DEFINITION M2-GN0127-231000-004-e11 GN0127 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF989696
VERSION BF989696.1 GI:12396021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongsomjit, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-27049922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&t2=MR2-GN0127-231000-004-e11&t3=2000-10-23&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.
Location/Qualifiers
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0127"
/dev_stage="Adult"
/note="Organ: placenta, normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORSTS PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 86 a 54 c 82 g 68 t 1 others

ORIGIN
Query Match 50.5%; Score 253.2; DB 12; Length 291;
Best Local Similarity 96.4%; Pred. No. 6.1e-70;
Matches 269; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 4 GATCTCAGATTATCATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCTTAT 63
Db 280 GAACCTAGGATATATCATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCTTAT 221
Qy 64 ACTGCTCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
Db 220 ACTGCTCTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161
Qy 124 GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 183
Db 160 GCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 101
Qy 184 TCAATCCCAATCTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 243
Db 100 TCAATCCCAATCTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 41
Qy 244 CATCTATTTGGCGAGGAGCTTACGCGAGCTTGAACCA 282
Db 40 CATCTATTT-GCGAGGAGCTTACGCGAGCTTGAACCA 3

RESULT 9

LOCUS B66771 707 bp DNA linear GSS 21-JUN-1998
DEFINITION CIT-HSP-2015D21.TFB CIT-HSP Homo sapiens genomic clone 2015D21, DNA
sequence.
ACCESSION B66771
VERSION B66771.1 GI:2640749
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 707)
AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Base,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CIT-HSP-2015D21.TR
CONTACT: Mark Adams
DEPARTMENT: Department of Eukaryotic Genomics
INSTITUTION: The Institute for Genomic Research
ADDRESS: 9712 Medical Center Dr., Rockville, MD 20850, USA
TEL: 301 838 0200
FAX: 301 838 0208
EMAIL: mdadams@tigr.org
CLONES ARE AVAILABLE FROM: Research Genetics (info@resgen.com). BAC
END SEARCH PAGE:
HTTP://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21
CLASS: BAC ends.

FEATURES
source

1..707
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7043039"
/db_xref="taxon:9606"
/clone="2015D21"
/clone_1id="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; site_1: HindIII; site_2:
HindIII"

BASE COUNT 195 a 183 c 144 g 185 t
ORIGIN

Query Match 50.4%; Score 252.6; 1B 17; Length 707;
Best Local Similarity 71.4%; Pred. No. 1.5e-69;
Matches 362; Conservative 0; Mismatches 139; Indels 6; Gaps 2;
1 CAAATCTCAGATTATCATGAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCCT 60
|||||

Db 176 CAAATCTCAGATTATCATGAGGCTGTGTCTCTATATGCCAGCTGTACCAACCC 235
Qy 61 TATATCTGCTTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 236 TATATCTGCTTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295
Qy 121 GATGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db 296 GATGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 355
Qy 181 ACTTCAATCCCAATCTCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 240
Db 356 CCGGAGAACCCATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 415
Qy 241 CCGCATCTATTTGGCGAGGAGCTTACGCGAGCTTGAACCACTTGAACCTCACTCACT 298
Db 416 CCGCATCTATTTGGCGAGGAGCTTACGCGAGCTTGAACCACTTGAACCTCACTCACT 475
Qy 299 CTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358
Db 476 CTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 535
Qy 359 CAAGCCACCCAGGAGGCTCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
Db 536 CAGGCTACTCTGATCTCTTGAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595
Qy 419 AAGGCTCACTCTGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 474
Db 596 AAGGCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 655
Qy 475 AAGGCCCTCACTGAGGAGACATCTCAG 501
Db 656 ATGCTCTTCACTCACTGAGGAGATATCAG 682

RESULT 10

LOCUS AV731083 736 bp mRNA linear EST 17-OCT-2000
DEFINITION AV731083 HTF Homo sapiens CDNA clone HTFAH01 5', mRNA sequence.
ACCESSION AV731083
VERSION AV731083.1 GI:10840504
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 736)
AUTHORS Gu,Y., Peng,Y., Song,H., Huang,Q., Yang,Y., Gao,G., Xiao,H., Xu,X.,
Li,N., Qian,B., Liu,F., Qu,J., Gao,X., Cheng,Z., Xu,Z., Zeng,L., Xu
,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Hu,R.,
Chen,J., Chen,Z., and Han,Z.
TITLE Homo sapiens CDNA HTF clones
JOURNAL Unpublished (2000)
CONTACT: Zeguang Han
INSTITUTION: Chinese National Human Genome Center at Shanghai
ADDRESS: 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
TEL: 86-21-50801919 (ex.45)
FAX: 86-21-50801922
EMAIL: hanzg@chgc.sh.cn
CLONE IS AVAILABLE AT: CHGC in Shanghai.

FEATURES
source

1..736
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTFAH01"
/clone_1id="HTF"
/issue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 193 a 147 c 185 g 211 t

ORIGIN

Query Match 41.5%; Score 208; DB 10; Length 736;
Best Local Similarity 74.9%; Pred. No. 2.9e-55;
Matches 274; Conservative 0; Mismatches 90; Indels 2; Gaps 1;

QY 1 CAAGATCTCAGATATATCATGAGGCTGTGTCTCTATAGCCAGCTGTACAGCCCT 60
DB 376 CAAGATCTTACATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 317
QY 61 TATATCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 316 TATACCTGCTCTCTCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 257
QY 121 GATGCTCTCTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 180
DB 256 GATGCTGCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 197
QY 181 ACTTCAACCAACATCTCAATCTCACTGATCTATTTTACCCAGAGGTTGAGGATAGT 240
DB 196 CCCACAGACACATCATCACTTATCTTGATGCTTGCCCAAGGTTTATAGGATAGC 137
QY 241 CCCATCTATTTGGCCAGGAGGATGAGCCCAAGAGCTTGAGCAATCTCTCA--TACCTGAGCA 298
DB 136 CCTCATCTGTTGGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 77
QY 299 CTTGCTCTGCTGATGAGTATGATGATTTACTTTTGGCCGCCAATTGAGAACTTGTCAT 358
DB 76 CTGGTCTCTGATGATGATGATGATTTACTTAATGCTTACATTTCAAGAAACATGCCCC 17
QY 359 CAAGCC 364
DB 16 CGTGCC 11

RESULT 11 537 bp DNA linear GSS 03-NOV-1998
AQ270891/c
LOCUS HS_2047_A1_H02_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION sapiens genomic clone Plate=2047 Col=3 Row=O, DNA sequence.
ACCESSION AQ270891
VERSION AQ270891.1 GI:3823431
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 537)
Mahaits G.G., Wallace J.C., Smith K., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahaits G.G., Wallace J.C., Hood L.
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2047 Row: O Column: 3
Class: BAC ends
High quality sequence stop: 537.
Location/Qualifiers
1..537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2047 Col=3 Row=O"
/clone_id="CIT Approved Human Genomic Sperm Library D"
/sex="male"

FEATURES
source

BASE COUNT 125 a 97 c 146 g 161 t 8 others
ORIGIN

Query Match 41.2%; Score 206.2; DB 17; Length 537;
Best Local Similarity 90.0%; Pred. No. 9.2e-55;
Matches 243; Conservative 0; Mismatches 21; Indels 6; Gaps 2;

QY 238 AGTCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 297
DB 536 AGCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 477
QY 298 A--CTTGCTCTGCTGATGATGATTTACTTTTGGCCGCCAATTGAGAACTTGTC 355
DB 476 ACTTGCTCTGCTGATGATGATTTACTTTTGGCCGCCAATTGAGAACTTGTC 417
QY 356 CATCAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAA 415
DB 416 CATCAAGCCACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 357
QY 416 CCAGAGGCTCACTCTGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 471
DB 356 CCAGAGGCTCACTCTGCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 297
QY 472 ACCAGGCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 501
DB 296 ACCAGGCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 267

RESULT 12 416 bp DNA linear GSS 14-JUL-1998
AQ041889/c
LOCUS CIT-HSP-2335P3.TR CIT-HSP Homo sapiens genomic clone 2335P3, DNA
DEFINITION sequence.
ACCESSION AQ041889
VERSION AQ041889.1 GI:3311083
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
Adams M.D., Rounsley S.D., Zhao S., Field C.E., Bass S., Linher K.,
Simon M., and Venter J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
CONTACT: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2335P3"
/clone_id="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11, Site_1: HindIII, Site_2:
HindIII"

FEATURES
source

BASE COUNT 119 a 85 c 113 g 99 t
ORIGIN

Seq primer: puc 18 forward
High quality sequence stop: 432.
Location/Qualifiers
1. 434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1id="HT0513"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 125 a 92 c 120 g 97 t
ORIGIN

Query Match 40.7%; Score 203.8; DB 10; Length 434;
Best Local Similarity 73.3%; Pred. No. 4.8e-54;
Matches 288; Conservative 0; Mismatches 102; Indels 3; Gaps 2;

Oy 1 CAAGATCTCAGGATATTAATGAGGCTGTGTTCTCTATAGCCAGCTGTACTTGCCCT 60
Db 413 CAAGATCTTAACTCATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 354
Oy 61 TATATCTGCTTCCCAATATCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Db 353 G-TACTGCTCTCTCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 295
Oy 121 GATGCTTCTTCTGCACTCCGTGACATCTGACATCTGACATCTGACATCTGACATCT 180
Db 294 GATGCTTCTTCTGATTCACCTGACATCTGACATCTGACATCTGACATCTGACATCT 235
Oy 181 ACTTAACCCCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 240
Db 234 CCCACAGACACATCTCCCACTTACGTCAGTCCTGCCCAGGAGGAGGAGGAGGAGG 175
Oy 241 CCCCATCTATTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 298
Db 174 CTTACCTCTTGTGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 115
Oy 299 CTTGCTCTTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 114 CTGGCTCTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 55
Oy 359 CAAGCACCACCAAGGCTCTTCAATTTCTCGCT 391
Db 54 CAGGCTACTTAGATCTCTGAACTTCTAGCT 22

RESULT 15 408 bp DNA linear GSS 02-JUL-1998
LOCUS A0033042/c
DEFINITION HS 2226 A1 D12 MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2226 Col=23 Row=G, DNA sequence.
ACCESSION A0033042
VERSION A0033042.1 GI:3285230
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 408)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2226 row: G column: 23
Class: BAC ends
High quality sequence stop: 408.
Location/Qualifiers
1. 408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1id="Plate=2226 Col=23 Row=G"
/sex="Male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 115 a 83 c 105 g 98 t 7 others
ORIGIN

Query Match 40.1%; Score 201; DB 17; Length 408;
Best Local Similarity 69.5%; Pred. No. 3.7e-53;
Matches 282; Conservative 0; Mismatches 122; Indels 2; Gaps 1;

Oy 43 CCAGCTGATCCTGATCTGCTTATCTGCTTCCCAATATCCAGAGGAGGAGGAGGAGGAGG 102
Db 408 CCAAGTGTACCAACTATATACATGCTCTCTCAATATCCAGAGGAGGAGGAGGAGGAGG 349
Oy 103 ACAGTCTGACCTTCAAGATGCTTCTTCTGATCCCTGATCCCTGATCCCTGATCCCT 162
Db 348 ACTGTTTGAACNTGAAGATGCTTCTTGAATGATCCATGACATGCTGCTGCTGCTGCT 289
Oy 163 TTGTTGCTTTGAAGATCTTCAACCAACATCTCACTCACTCACTCACTCACTCACTCA 222
Db 288 CTTTGGCTTTGAGGATCCACAGACCAACGTCACATTTATATGATGACGTTTGGCT 229
Oy 223 CAAGGTTTCAAGGATGATCCCATCTATTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 282
Db 228 CAAAGGTTTGAAGATGATCCCATCTGTTGGTCAAGATATGAGGAGGAGGAGGAGGAG 169
Oy 283 TCTCTCA--TACCTGAGACATTTGCTTGGTATGATGATGATGATGATGATGATGATG 340
Db 168 TTCTCAAGTCCAGGACATCTGCTCTTCAATATGATGATGATGATGATGATGATGATG 109
Oy 341 TCAAGAACCTTGTGCTCAAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 400
Db 108 TGAAGCTTTCATGCTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 49
Oy 401 TACATGTTTCCAAACCAAGGCTCAACTGCTCAGAGGAGGAGGAGGAGGAGGAGGAGG 446
Db 48 TACAAAGCATTAACCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3

Search completed: April 17, 2003, 07:35:31
Job time : 1275.84 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:07:23 ; Search time 1700.07 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Sequence: 1 cccctggggcggccttcctt.....tgactcccatccctcgtgat 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenBml:.*
1: gb_ba:.*
2: gb_htg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_stg:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_stg:.*
28: em_un:.*
29: em_vi:.*
30: em_htg_hum:.*
31: em_htg_inv:.*
32: em_htg_other:.*
33: em_htg_mus:.*
34: em_htg_pln:.*
35: em_htg_rtd:.*
36: em_htg_mam:.*
37: em_htg_vtc:.*
38: em_sy:.*
39: em_htgo_hum:.*
40: em_htgo_mus:.*
41: em_htgo_other:.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	10499	6 AX007980	AX007980 Sequence
2	500	100.0	56093	6 AX329572	AX329572 Sequence
3	500	100.0	56093	9 HSA0000064	AC0000064 Human BAC
4	500	100.0	149194	9 AC007566	AC007566 Homo sapi
5	435	87.0	711	6 AX007997	AX007997 Sequence
6	410	82.0	783	6 AX000970	AX000970 Sequence
7	383.8	76.8	711	6 AX007998	AX007998 Sequence
8	360	72.0	137492	2 AL161721	AL161721 Homo sapi
9	359.8	72.0	143590	2 HSB3613	AL161721 Homo sapi
10	358.2	71.6	119406	2 AC103595	AC103595 Homo sapi
11	357.8	71.6	110000	2 AL353584_0	AL353584 Homo sapi
12	357.8	71.6	169029	9 CENS06C7R	AL353584 Homo sapi
13	357.8	71.6	170746	9 AL392173	AL392173 Human chr
14	357.4	71.5	163803	9 AC093531	AC093531 Homo sapi
15	357.4	71.5	191863	9 AC010888	AC010888 Homo sapi
16	356.8	71.4	180635	9 AL360169	AL360169 Homo sapi
17	356.6	71.3	174019	9 AP001538	AP001538 Homo sapi
18	356.2	70.8	340000	9 AP001674	AP001674 Homo sapi
19	354.2	70.8	111079	9 HS419C19	AL035407 Human DNA
20	353.4	70.7	76169	9 AC003014	AC003014 Human DNA
21	351.8	70.4	161049	2 AL591842	AL591842 Homo sapi
22	351	70.2	42216	9 AC092899	AC092899 Homo sapi
23	351	70.2	170414	2 AC026019	AC026019 Homo sapi
24	349.6	69.9	83412	9 AC092843	AC092843 Homo sapi
25	349.2	69.6	165059	9 AC106856	AC106856 Homo sapi
26	348.2	69.6	839	9 AF127226	AF127226 Homo sapi
27	348.2	69.6	1329	6 AX001030	AX001030 Sequence
28	347.6	69.5	177212	2 AC025723	AC025723 Homo sapi
29	347.6	69.5	181753	9 AC124075	AC124075 Homo sapi
30	347.2	69.4	119481	9 AL136234	AL136234 Human DNA
31	347.2	69.4	152996	9 AC013719	AC013719 Homo sapi
32	346.4	69.3	179585	2 AC024977	AC024977 Homo sapi
33	346.4	69.3	185671	9 AC084193	AC084193 Homo sapi
34	346.2	69.2	156177	9 AC093896	AC093896 Homo sapi
35	345	69.0	37764	9 HS0221F2	275746 Human DNA s
36	345	69.0	143620	9 AC117425	AC117425 Homo sapi
37	344.8	69.0	157711	9 AC055738	AC055738 Homo sapi
38	344.8	69.0	181850	9 AC018680	AC018680 Homo sapi
39	344.8	69.0	207945	2 CENS01DSS	AL121841 Human chr
40	344.8	69.0	318442	2 AC021652	AC021652 Homo sapi
41	344.6	68.9	140756	9 AL139038	AL139038 Human DNA
42	344.6	68.9	183499	2 AL607153	AL607153 Homo sapi
43	344.4	68.9	88328	9 AL357874	AL357874 Human DNA
44	343	68.6	187321	2 AC092510	AC092510 Papio cyn
45	341.6	68.3	102200	9 HS30M3	AL031775 Human DNA

ALIGNMENTS

```
RESULT 1
AX007980
LOCUS AX007980 10499 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO967395.
ACCESSION AX007980
VERSION AX007980.1 GI:9995677
KEYWORDS
SOURCE
ORGANISM human.
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 10499)
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
```

JOURNAL Patent: WO 967395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERLIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALIET, PATRICK M (FR)
FEATURES
Source Location/Qualifiers
1. 10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8,5e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCTGGGGGGGGCTTCCTTTCTGGATGAGGGCAAAACGCTCGAGATACAGCAATTATC 60
DB 1 CCTGGGGGGGGCTTCCTTTCTGGATGAGGGCAAAACGCTCGAGATACAGCAATTATC 60
QY 61 TTGCACTGAGAGACAGGACTAGCTGATTTCTTAGACCGACTAAGAAATCCCTAAGCCTA 120
DB 61 TTGCACTGAGAGACAGGACTAGCTGATTTCTTAGACCGACTAAGAAATCCCTAAGCCTA 120
QY 121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGAATTAGCTCACACTGA 180
DB 121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGAATTAGCTCACACTGA 180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGGATAAGAAATAGCCA 240
DB 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGGATAAGAAATAGCCA 240
QY 241 ATCATCTATTGCTGAGACACAGAGGAGGAGCAACATCGGATATATAACCCAGGGCAT 300
DB 241 ATCATCTATTGCTGAGACACAGAGGAGGAGCAACATCGGATATATAACCCAGGGCAT 300
QY 301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCTTTGATGGAGAGCTTTTCA 360
DB 301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCTTTGATGGAGAGCTTTTCA 360
QY 361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCTCTGCTCATGTTTCTTACGGCT 420
DB 361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCTCTGCTCATGTTTCTTACGGCT 420
QY 421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTCGCGC 480
DB 421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTCGCGC 480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 481 TGACTCCCATCCCTCTGGAT 500
```

RESULT 2
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 81 from Patent WO019629.
DEFINITION AX329572
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G.,
Horrikan, S., Soper, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
Source Location/Qualifiers
1. 56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 56093;
Best Local Similarity 100.0%; Pred. No. 9,4e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCTGGGGGGGGCTTCCTTTCTGGATGAGGGCAAAACGCTCGAGATACAGCAATTATC 60
DB 28001 CCTGGGGGGGGCTTCCTTTCTGGATGAGGGCAAAACGCTCGAGATACAGCAATTATC 28060
QY 61 TTGCACTGAGAGACAGGACTAGCTGATTTCTTAGACCGACTAAGAAATCCCTAAGCCTA 120
DB 28061 TTGCACTGAGAGACAGGACTAGCTGATTTCTTAGACCGACTAAGAAATCCCTAAGCCTA 120
QY 121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGAATTAGCTCACACTGA 180
DB 28121 GCTGGGAAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGAATTAGCTCACACTGA 28180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGGATAAGAAATAGCCA 240
DB 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGCAGAGGATAAGAAATAGCCA 28240
QY 241 ATCATCTATTGCTGAGACACAGAGGAGGAGCAACATCGGATATATAACCCAGGGCAT 300
DB 28241 ATCATCTATTGCTGAGACACAGAGGAGGAGCAACATCGGATATATAACCCAGGGCAT 28300
QY 301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCTTTGATGGAGAGCTTTTCA 360
DB 28301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCTTTGATGGAGAGCTTTTCA 28360
QY 361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCTCTGCTCATGTTTCTTACGGCT 420
DB 28361 TGCATTTTCACTCTATTAAATCTTGCAACTGCACTCTCTGCTCATGTTTCTTACGGCT 28420
QY 421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTCGCGC 480
DB 28421 CGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTT GCCACACCGCAGACTCGCGC 28480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 28481 TGACTCCCATCCCTCTGGAT 28500
```

RESULT 3
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
LOCUS Human BAC clone RG083M05 from 7q21-7q22, complete sequence.
DEFINITION AC000064
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS
SOURCE HTG.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 56093)
AUTHORS Pauley, A.
TITLE The sequence of H. sapiens BAC clone RG083M05
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 56093)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBELO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H_RG083M05; actual end is at 56093 of H_RG083M05

This clone contains SRS SMS51725.

FEATURES

source

Location/Qualifiers

1..56093

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="H_RG083M05"

/clone_1lb="CITB-978SK-B"

/complement(838..1131)

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(1360..1503,4181..4370,4587..4774,6422..6556,9483..9547,11631..11773,11864..12021,13131..13296,14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="ATPase; strong similarity to peroxisome biosynthesis protein PAB1 (PID:g117019); coded for by

human cDNA C04279 (NID:g1467550)"

/codon_start=1

/protein_id="AAB6346.1"

/db_xref="GI:1669371"

/translation="KRLNLTQKTEVAESEAVMNOPSVLLDLDLIAGLPAVBEHEHSPBARCEILCNVINKLDCDINKFTDLDLGVARETGFVARDPTVADRAIHERLSRSISITREKLVTTIDPQKALRGFPASLRSSNUNKPRDLGDKIGGHEVQIILMDTLOPAKVCILKKEKYPELFANLPIRQRTGILVPGPGTKTLGAVIARESRNFTSV

KEBELSKYIGASEQAVRDIPIRAQAKPCILFPDEFESIAPRRHDNVTGVVNO

LLTQLDGEGLQGVYLAATSRDLDIPALIRGRIDKCYCPPOVITISYESTQ

OMHSFLVSRLEILNVLSDELPLADDVDLQHVASVDSFGADLKALLVVAOLEALHG

MLSKMSKELIPDSKFMVRLYFGSSYSESLGNGTSLSOCLSA PSSWTDLPGVP

GKQULFSQPVYLTASQEGCELTQGEORDDLRADISLIKRYRNSQGEDESNMOPPI

KTRLAISQSHLMALGHTRPSISEDWKNFAEL"

/complement(4948..5130)

/rpt_family="ALU"

/complement(6581..7133)

/rpt_family="U1"

/complement(7767..8037)

/rpt_family="ALU"

/complement(8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:g1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:g1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N22627 (NID:g1130501), bases

276-343"

misc_feature

repeat_region

complement(12612..12907)

/rpt_family="ALU"

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

143-266"

repeat_region

13794..13877

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g917434), bases

30-58"

repeat_region

13907..14104

/rpt_family="ALU"

complement(14110..14137)

/rpt_family="U1"

complement(15618..15907)

/rpt_family="ALU"

17227..17522

/rpt_family="ALU"

18667..19235

/note="match to human fetal brain 5' EST D61494

(NID:g970409), bases 1-255, and to human 3' EST R07476

(NID:g759399)"

19550..19670

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polypeptides; pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:g942257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

complement(138538..139224)

/rpt_family="ALU"

39225..39707

/note="match to multiple human ESTs, see N30113

(NID:g1148633)"

39800..40085

/rpt_family="ALU"

complement(40247..40538)

/rpt_family="ALU"

complement(40632..40924)

/rpt_family="ALU"

complement(42283..42891)

/rpt_family="ALU"

complement(45474..45613)

/rpt_family="ALU"

complement(45614..45737)

/note="match to human 3' EST H48898 (NID:g988738), bases

129-333"

misc_feature

complement(46107..47026)

/note="match to multiple human ESTs, see N81064

(NID:g1243765), H48897 (NID:g988737), and M78831

(NID:g273146)"

complement(47027..47318)

/rpt_family="ALU"

complement(47365..47782)

/note="match to multiple human ESTs, see W37495

(NID:g1319089)"

47898..48115

/note="match to human 5' EST H62306 (NID:g1015138), bases

93-368"

repeat_region

complement(48116..48405)

/rpt_family="ALU"

complement(48406..48584)

/note="match to human 3' EST N29952 (NID:g1148472), bases

250-455, and 5' EST R12730 (NID:g765806)"

complement(48787..49405)

/rpt_family="ALU"

complement(49406..49534)

/note="match to human 3' EST R65794 (NID:g838432), bases

309-440"

misc_feature

```

repeat_region complement(49638..49672)
               /rpt_family="ALU"
misc_feature complement(49674..49890)
               /note="match to human 3' EST N29952 (NID:g1148472) and 5'
               EST N29938 (NID:g1148450), sequences are from opposite
               ends of the same clone"
gene complement(49698..51806)
               /gene="WUGSC:R.RG083W05.2"
               /complement(join(49698..49888,51575..51806))
               /gene="WUGSC:R.RG083W05.2"
               /note="WUGSC:R.RG083W05.2"
               /note="coded for by human CDNA W37389 (NID:g119205),
               R65891 (NID:g838529), R65794 (NID:g838432) and R65794
               (NID:g838432)"
               /protein_id="AAB46345.1"
               /codon_start=1
               /translation="MPFYRQCGTTFPCPGVYVQIDVYVIDEDQKPYAIGRF
               IQDYCKSAALTLIPLTSPRD)PDPASTITGPEEDLPKMEYLEFVCHAPSEYRK
               SRSPFPVTPRPKGYIWHVGP:PAITIKESVANHL"
               /complement(51576..51758)
               /gene="WUGSC:R.RG083W05.2"
               /note="Grail prediction, score = 86"
               /evidence=not_experimental
repeat_region complement(52052..52329)
               /rpt_family="L1"
               /note="match to human EST M79192 (NID:g273505) base 2-289"
misc_feature 55557..55843
Query Match 100.0%; Score 500; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 9,4e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTGTGGGCGGGCTTCCTTTCTGGATGAGGCAAAACGCTCGAGATACAGCAATATC 60
Db 28001 CCTGTGGGCGGGCTTCCTTTCTGGATGAGGCAAAACGCTCGAGATACAGCAATATC 28060
QY 61 TTGCAACTGAGAGACAGAGACTAGTGGATTCTTACGCGACTAAGAAATCCCTAAGCCTA 120
Db 28061 TTGCAACTGAGAGACAGAGACTAGTGGATTCTTACGCGACTAAGAAATCCCTAAGCCTA 28120
QY 121 GCTGGGAAGGTGACCAACGCTTCAACCGGGGTTTGCACTTAGCTACACCTTGA 180
Db 28121 GCTGGGAAGGTGACCAACGCTTCAACCGGGGTTTGCACTTAGCTACACCTTGA 28180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAAATTAGCAAAACAGAGAGTAAAGAAATAGCCA 240
Db 28181 CCAATCAGAGAGCTCACTAAATGCTAAATTAGCAAAACAGAGAGTAAAGAAATAGCCA 28240
QY 241 ATCATCTATTGCTTGAAGACAGACAGAGAGGAGCAACATCGGGATATTAACCCAGGCAT 300
Db 28241 ATCATCTATTGCTTGAAGACAGACAGAGAGGAGCAACATCGGGATATTAACCCAGGCAT 28300
QY 301 TCGAGCTGGCAACACAGACCCCTTTGGGTCCCTTCTTGATTTGGAGCTGTTTTCA 360
Db 28301 TCGAGCTGGCAACACAGACCCCTTTGGGTCCCTTCTTGATTTGGAGCTGTTTTCA 28360
QY 361 TGCATTTCACTCTTAAATCTTGCAACTGCACTCTTGTGCTCATGTTTTCTTAGGCT 420
Db 28361 TGCATTTCACTCTTAAATCTTGCAACTGCACTCTTGTGCTCATGTTTTCTTAGGCT 28420
QY 421 CGAGCTGAGCTTTTCTCAGCGTCCACCACTGCTGTTTTCACCAACCGAGACTGCCGC 480
Db 28421 CGAGCTGAGCTTTTCTCAGCGTCCACCACTGCTGTTTTCACCAACCGAGACTGCCGC 28480
QY 481 TGACTCCCATCCCTCTGGAT 500
Db 28481 TGACTCCCATCCCTCTGGAT 28500

```

```

RESULT 4
AC007566/c AC007566 149194 bp DN\ linear PRI 01-MAR-2002
LOCUS Homo sapiens BAC clone CTB-10G5 from 'g21-7g22', complete sequence.
DEFINITION

```

```

ACCESSION AC007566
VERSION AC007566.2 GI:11181861
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
          1 (bases 1 to 149194)
          Sulston, J.E. and Waterston, R.
          Toward a complete human genome sequence
          Genome Res. 8 (11), 1097-1108 (1998)
          98063792
          9847074
MEDLINE PUBMED
REFERENCE 2 (bases 1 to 149194)
          Du, Z.
          The sequence of Homo sapiens BAC clone CTB-10G5
          Unpublished (2001)
          3 (bases 1 to 149194)
          Waterston, R.H.
          Direct Submission
          Submitted (15-MAY-1999) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          4 (bases 1 to 149194)
          Waterston, R.
          Direct Submission
          Submitted (02-OCT-2000) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          5 (bases 1 to 149194)
          Waterston, R.H.
          Direct Submission
          Submitted (16-NOV-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          6 (bases 1 to 149194)
          Waterston, R.H.
          Direct Submission
          Submitted (03-JAN-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          7 (bases 1 to 149194)
          Waterston, R.H.
          Direct Submission
          Submitted (06-FEB-2002) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
          8 (bases 1 to 149194)
          Waterston, R.
          Direct Submission
          Submitted (01-MAR-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT ----- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: sapliens@wuston.wustl.edu
          ----- Summary Statistics
          -----
          Center project name: H_RG010G05

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:sgreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-9768SK-B. The library contains cloned DNA from the male fibroblast cell line 9768SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```
1. 149194
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="7"
   /map="7q21-7q22"
   /clone_1b="CTB-10G5"
   /clone_1b="CTB-9768SK-B"
   /rpt_family="L1"
   /rpt_family="L1"
   /note="match to EST BG752883 (NID:g14063536) "
   /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
   /note="match to EST AM579261 (NID:g7254310) "
   /note="match to EST BG766882 (NID:g14077535) "
   /note="match to EST B1160365 (NID:g14620366) "
   /note="match to EST B1160365 (NID:g14620366) "
   /note="similar to Homo sapiens EST BF758865 (NID:g12106765) "
   /note="match to EST AV686676 (NID:g10288539) "
   /note="match to EST AU123510 (NID:g10948226) "
   /note="match to EST AV686676 (NID:g10288539) "
   /note="match to EST AV686676 (NID:g10288539) "
   /rpt_family="Alu"
   /rpt_family="Alu"
   /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
   /note="similar to Mus musculus EST BB253526 (NID:g8946272) "
   /note="match to EST AV686676 (NID:g10288539) "
   /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
   /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
   /note="match to EST AU123510 (NID:g10948226) "
   /note="match to EST BG766882 (NID:g14077535) "
   /note="match to EST B1160365 (NID:g14620366) "
   /note="similar to Mus musculus EST BE994936 (NID:g14077535) "
```

```
misc_feature 3542..3628
              /note="match to EST B1160365 (NID:g14620366) "
misc_feature 3542..3609
              /note="match to EST AM579261 (NID:g7254310) "
misc_feature 3562..3628
              /note="match to EST BE272564 (NID:g9146913) "
misc_feature 3598..3628
              /note="match to EST BE299708 (NID:g9183456) "
misc_feature 3714..3785
              /note="match to EST BG260659 (NID:g12770475) "
misc_feature 3717..3785
              /note="similar to Mus musculus EST BE994936 (NID:g10678674) "
misc_feature 3719..3785
              /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
              /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature 3719..3785
              /note="match to EST AU123510 (NID:g10948226) "
misc_feature 3719..3785
              /note="match to EST BE272564 (NID:g9146913) "
misc_feature 3719..3785
              /note="match to EST BE299708 (NID:g9183456) "
misc_feature 3719..3785
              /note="match to EST BG766882 (NID:g14077535) "
misc_feature 3719..3785
              /note="match to EST B1160365 (NID:g14620366) "
misc_feature 3719..3785
              /note="similar to Mus musculus EST BB253526 (NID:g8946272) "
misc_feature 3860..3862
              /note="match to EST AV686676 (NID:g10288539) "
repeat_region 4730..4827
              /rpt_family="L2"
              /rpt_family="L2"
              /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
repeat_region 4834..4925
              /rpt_family="L1"
              /note="match to EST AU123510 (NID:g10948226) "
misc_feature 5715..5860
              /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature 5715..5860
              /note="match to EST AU123510 (NID:g10948226) "
misc_feature 5715..5860
              /note="match to EST BE272564 (NID:g9146913) "
misc_feature 5715..5860
              /note="match to EST BE299708 (NID:g9183456) "
misc_feature 5715..5860
              /note="match to EST BG260659 (NID:g12770475) "
misc_feature 5715..5860
              /note="match to EST BG766882 (NID:g14077535) "
misc_feature 5715..5860
              /note="match to EST B1160365 (NID:g14620366) "
misc_feature 5715..5860
              /note="similar to Mus musculus EST BE994936 (NID:g14077535) "
```

Query Match 100.0%; Score 500; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 1e-154;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCTGGGCGGGCTCTCTTCTGGATGAGGCAAAACGCTTGAAGTATAGCAATTATC 60
DB 94003 CCTGGGCGGGCTCTCTTCTGGATGAGGCAAAACGCTTGAAGTATAGCAATTATC 93944
QY 61 TTGCACTGAGAGACAGACTAGTGGATTCTTCTAGGCGAGCTTAAGATCCCTAAGCCTA 120
DB 93943 TTGCACTGAGAGACAGACTAGTGGATTCTTCTAGGCGAGCTTAAGATCCCTAAGCCTA 93884
QY 121 GCTGGAGAGGTGACCACTGACCTTTAAACAGGCGGCTTGAACCTTAGCTCACCTGA 180
DB 93883 GCTGGAGAGGTGACCACTGACCTTTAAACAGGCGGCTTGAACCTTAGCTCACCTGA 93824
QY 181 CCAATCAGAGAGCTCACTAATAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCA 240
DB 93823 CCAATCAGAGAGCTCACTAATAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCA 93764
```

QY 241 ATCATCTATTGCTGAGAGACAGAGAGAGACACAAATCGGGATATTAACCCAGGCAT 300
Db 93763 ATCATCTATTGCTGAGAGACAGAGAGAGAGACACAAATCGGGATATTAACCCAGGCAT 93704
QY 301 TCGAGCTGGCAACAGACAGACCCCTTTGGGTCCTTCCCTTTGTAATGGAGCTGTTTCA 360
Db 93703 TCGAGCTGGCAACAGACAGACCCCTTTGGGTCCTTCCCTTTGTAATGGAGCTGTTTCA 93644
QY 361 TGCATTTTCACTATTAAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCT 420
Db 93643 TGCATTTTCACTATTAAATCTTGAACCTGCACTTTCTGGTCCATGTTTCTTAAGGCT 93584
QY 421 CGAGCTGAGCTTTTGTGCTCACCGTCCACCACTGCTGTTTGCCACCAACCCGACAGCTGCC 480
Db 93583 CGAGCTGAGCTTTTGTGCTCACCGTCCACCACTGCTGTTTGCCACCAACCCGACAGCTGCC 93524
QY 481 TGACTCCCATCCCTCTGGAT 500
Db 93523 TGACTCCCATCCCTCTGGAT 93504

RESULT 5
AX007997 711 bp DNA linear PAT 06-SEP-2000
LOCUS AX007997 Sequence 20 from Patent WO9667395.
DEFINITION AX007997
ACCESSION AX007997
VERSION AX007997.1 GI:995694
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 711)
AUTHORS Perin, J.P., Rieger, F. and Alliel, P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL Patent: WO 9667395-A 20 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
source 1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 170 a 204 c 162 g 175 t
ORIGIN

Query Match 87.0%; Score 435; DB 6; Length 711;
Best Local Similarity 100.0%; Pred. No. 3,76-133; Indels 0; Gaps 0;
Matches 435; Conservative 0; Mismatches 0;

QY 66 ACTGAGAGACAGAGCTAGCTGATTTCTAGAGCCGACTAAGAATCCCTAAGCCTAGCTGG 125
Db 1 ACTGAGAGACAGAGCTAGCTGATTTCTAGAGCCGACTAAGAATCCCTAAGCCTAGCTGG 60
QY 126 GAAGGTGACAGCTGCACTTTAAACACGGGCTTGCACTTAAGTCCACCTGACCAAT 185
Db 61 GAAGGTGACAGCTGCACTTTAAACACGGGCTTGCACTTAAGTCCACCTGACCAAT 120
QY 186 CAGAGAGCTCAATAAATCTAATTAGGCAAGAGAGAGAGTAAGTAAGTAAGCAATCAT 245
Db 121 CAGAGAGCTCAATAAATCTAATTAGGCAAGAGAGAGTAAGTAAGTAAGCAATCAT 180
QY 246 CTATTGCTGAGAGACAGACAGAGAGAGCAACAATCGGATATAAACCAGGACTTCGAG 305
Db 181 CTATTGCTGAGAGACAGACAGAGAGAGCAACAATCGGATATAAACCAGGACTTCGAG 240
QY 306 CTGGCAACAGAGACCCCTTTGGGTCCCTTCCTTTGTAATGGAGCTTTTCAATGCTA 365
Db 241 CTGGCAACAGAGACCCCTTTGGGTCCCTTCCTTTGTAATGGAGCTTTTCAATGCTA 300
QY 366 TTTCACCTATTAAATCTTGGCACTGCTTCTGCTCAATGTTTCTTAAGGCTCGAGC 425
Db 301 TTTCACCTATTAAATCTTGGCACTGCTTCTGCTCAATGTTTCTTAAGGCTCGAGC 360

QY 426 TGAGCTTTTGTCAACCGTCCACCACTGCTGTTTGCCACCAACCCGACAGCTGCGCTACT 485
Db 361 TGAGCTTTTGTCAACCGTCCACCACTGCTGTTTGCCACCAACCCGACAGCTGCGCTACT 420
QY 486 CCCATCCCTCTGGAT 500
Db 421 CCCATCCCTCTGGAT 435

RESULT 6
AX000970 783 bp DNA linear PAT 10-MAR-2000
LOCUS AX000970 Sequence 15 from Patent WO902696.
DEFINITION AX000970
ACCESSION AX000970
VERSION AX000970.1 GI:7241212
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 783)
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 902696-A 15 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source 1..783
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 173 a 213 c 166 g 180 t 51 others
ORIGIN

Query Match 82.0%; Score 410; DB 6; Length 783;
Best Local Similarity 93.8%; Pred. No. 6,8e-125;
Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;

QY 68 TGAGAGACAGAGCTAGCTGATTTCTAGAGCCGACTAAGAATCCCTAAGCCTAGCTGG 127
Db 1 TGAGAGACAGAGCTAGCTGATTTCTAGAGCCGACTAAGAATCCCTAAGCCTAGCTGG 60
QY 128 AGGAGACAGAGCTGCACTTTAAACACGGGCTTGCACTTAAGTCCACCTGACCAATCA 187
Db 61 AGGAGACAGAGCTGCACTTTAAACACGGGCTTGCACTTAAGTCCACCTGACCAATCA 120
QY 188 GAGAGCTCAATAAATGCTAATTAGGCAAGAGAGAGTAAGTAAGTAAGCAATCATCT 247
Db 121 GAGAGCTCAATAAATGCTAATTAGGCAAGAGAGAGTAAGTAAGTAAGCAATCATCT 180
QY 248 ATTGCTGAGAGACAGAGAGAGAGCAACAATGGGATATTAACCCAGGCAATTCGAGCT 307
Db 181 ATTGCTGAGAGACAGAGAGAGAGCAACAATGGGATATTAACCCAGGCAATTCGAGCT 240
QY 308 GGCAACAGACAGACCCCTTTGGGTCCCTTCCTTTGTAATGGAGAG--CTGTTTCAATGCTA 365
Db 241 GGCAACAGACAGACCCCTTTGGGTCCCTTCCTTTGTAATGGAGAGCTGTTTCAATGCTA 300
QY 366 TTTCACCTATTAAATCTTGGCACTGCACTTTCTGCTCAATGTTTCTTAAGGCTCGAGC 425
Db 301 TTTCACCTATTAAATCTTGGCACTGCACTTTCTGCTCAATGTTTCTTAAGGCTCGAGC 360
QY 426 TGAGCTTTTGTCAACCGTCCACCACTGCTGTTTGCCACCAACCCGACAGCTGCGCTACT 485
Db 361 TGAGCTTTTGTCAACCGTCCACCACTGCTGTTTGCCACCAACCCGACAGCTGCGCTACT 420
QY 486 CCCATCCCTCTGGAT 500
Db 421 CCCATCCCTCTGGAT 435

RESULT 7
AX007998 711 bp DNA linear PAT 06-SEP-2000
LOCUS AX007998

DEFINITION Sequence 21 from Patent WO9967395.
ACCESSION AX007998
VERSION AX007998.1 GI:9995695
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 711)
AUTHORS
TITLE
JOURNAL
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
Patent: WO 9967395-A 21 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALIHEL PATRICK M (FR)
LOCATION/Qualifiers
1. 711
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 171 a 209 c 160 g 171 t
ORIGIN
Query Match 76.8%; Score 383.8; DB 6; Length 711;
Best Local Similarity 95.4%; Pident. No. 3.6e-116;
Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;
QY 66 ACTGAGAGACAGAGCTGCTGATTTCTTCTAGCCGACTAAGAAATCCCTAAGCCTAGCTGG 125
Db 1 ACTGAGAGACAGAGCTGCTGATTTCTTCTAGCCGACTAAGAAATCCCTAAGCCTAGCTGG 60
QY 126 GAAAGTACACACGCTCCCTTTAAACAGGAGGCTTGGCACTTACCTACACCTGACCAAT 185
Db 61 GAAAGTACACACCTCCCTTTAAACAGGAGGCTTGGCACTTACCTACACCTGACCAAT 120
QY 186 CAGAGAGCTACTTAAATGCTAATTAGCAAAAGCAGAGGTAAAGAAATAGCAATCAT 245
Db 121 CAGAGAGCTACTTAAATGCTAATTAGCAAAAGCAGAGGTAAAGAAATAGCAATCAT 180
QY 246 CTATGCTGAGAGCAGAGAGAGAGCAACAATCGGATATTAACCCAGGACTTGGAG 305
Db 181 CTATGCTGAGAGCAGAGAGAGAGCAACAATCGGATATTAACCCAGGACTTGGAG 240
QY 306 CTGCGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGTATGGAG--CTGTTTTCATGC 363
Db 241 CCGGCAACGCGA-AACCCCTTTGGGTCCTTCCCTTTGTATGGAGCTGTTTTCATGC 299
QY 364 TATTTCTCTATTAATCTTTGCACTGCACTCTTCTGTCATGTTTCTTAGCGCTCGA 423
Db 300 TATTTCTCTATTAATCTTTGCACTGCACTCTTCTGTCATGTTTCTTAGCGCTTGA 359
QY 424 GCTGAGCTTTGCTCAACCGTCCACCACTGCTGTTGCAACCGGAGAGCTGCGGCTGA 483
Db 360 GCTGAGCTTTGCTGCGCACTGCACTGCTGTTGCGCCACCGGAGAGCTGCGGCTGA 419
QY 484 CTCCCATCCCTCTGAT 500
Db 420 CTCCCATCCCTCTGAT 436
RESULT 8
AL161721/c 137492 bp DNA linear HTG 10-JUL-2001
LOCUS Homo sapiens chromosome 6 clone RPI-301F24 map q25.2-26, ***
DEFINITION SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION AL161721 GI:12330750
VERSION AL161721.3 HTG; HTGS_PHASE1; HTGS_CANCELLED.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 137492)
REFERENCE
AUTHORS
Stim,S.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Jan 22, 2001 this sequence version replaced gi:9212839.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dj301F24
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 13573 bases at least Q40
Consensus quality: 135245 bases at least Q30
Consensus quality: 136166 bases at least Q20
Insert size: 136992; sum-of-contigs
Insert size: 152637; 0.5% error; agarose-fp
Quality coverage: 3.96x in Q20 bases; sum-of-contigs Quality
coverage: 3.56x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12245: contig of 12245 bp in length
* 12246 12345: gap of 100 bp
* 12346 20428: contig of 8083 bp in length
* 20429 20528: gap of 100 bp
* 20529 85104: contig of 64576 bp in length
* 85105 85204: gap of 100 bp
* 85205 96812: contig of 11608 bp in length
* 96813 96912: gap of 100 bp
* 96913 112622: contig of 15710 bp in length
* 112623 112722: gap of 100 bp
* 112723 137492: contig of 24770 bp in length.
* Location/Qualifiers
1. 137492
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.2-26"
/clone_1db="RPI-1"
/clone_1db="RPI-1"
1. 12245
/note="assembly_fragment:00933
fragment_chain:1
clone_end:17
vector_side:left"
12346..20428
/note="assembly_fragment:01071
fragment_chain:1"
20529..85104
/note="assembly_fragment:00093
fragment_chain:1"
85205..96812
/note="assembly_fragment:00954
fragment_chain:1"
96913..112622
/note="assembly_fragment:01341"
112723..137492
/note="assembly_fragment:00790
clone_end:86
vector_side:right"
BASE COUNT 37669 a 28026 c 30367 g 40929 t 501 others
ORIGIN

Query Match 72.0%; Score 360; DB 2; Length 137492;
 Best Local Similarity 89.6%; Pred. No. 4,1e-108; Indels 1; Gaps 1;
 Matches 399; Conservative 0; Mismatches 45;

Qy 57 TATCTTGCACCTGAGAGACAGACTAGCTGATTTCTTACGCCGACCTAAGATCCCTAAG 116
 Db 128053 TTTAATGCTTGAAGAGACAGACTAGCTGATTTCTTACGCCGACCTAAGATCCCTAAG 127994
 Qy 117 CCTAGCTGGAGAGTGAACACAGCTTCACTTTAAACA-3GGGGCTTGCAACTTACCTCACAC 176
 Db 127993 CCTAGCTGGAGAGTGAACACAGCTTCACTTTAAACA-3GGGGCTTGCAACTTACCTCACAC 127934
 Qy 177 CTGACCATCAGAGAGCTCCTAATAATGCTAATTAG;CAAAAGACAGAGGTAAAGAAATA 236
 Db 127933 CCAACCATCAGAGAGCTCCTAATAATGCTAATTAG;CAAAAGACAGAGGTAAAGAAATA 127874
 Qy 237 GCCAATCATCTATTGCTGAGAGCAGACAGAGAGG;CAACATCGGATATTAACCCAG 296
 Db 127873 GCCAATCATCTATTGCTGAGAGCAGACAGAGAGG;CAACATCGGATATTAACCCAG 127814
 Qy 297 GCATTGAGCTGGCAACAGACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTT 356
 Db 127813 GCATTGAGCTGGCAACAGACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTT 127755
 Qy 357 TTCATGCTATTCTACTCTATTAAATCTTGCAACTGCTCTTGGTCCATGTTTCTTAC 416
 Db 127754 TTCACTATTCTACTCTATTAAATCTTGCAACTGCTCTTGGTCCATGTTTCTTAC 127695
 Qy 417 GGCTGAGCTGAGCTTTTGTCTCACCGTCCACCATGCTGTTTGCACACCGCAGACCTG 476
 Db 127694 GGCTGAGCTGAGCTTTTGTCTCACCGTCCACCATGCTGTTTGCACACCGCAGACCTG 127635
 Qy 477 CCGCTGATCTCCATCTCTCTGAT 500
 Db 127634 CCGCTGATCTCCATCTCTCTGAT 127611

RESULT 9
 HS836J3/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP5-836J3 on chromosome 1p22.1-31.1
 ACCESSION AL035706
 VERSION AL035706.10 GI:5777502
 KEYWORDS HTG; CPG Island.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 143590)
 Bagguley, C.
 Direct Submission
 Submitted (29-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Aug 26, 1999 this sequence version replaced gi:5785651.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at

FEATURES

source
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chrl
 RP5-836J3 is from the library RPL1-5 constructed at the Roswell
 Park Cancer Institute by the group of Pliet de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP5-836J3.
 Location/Qualifiers
 1..143590
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="p22.1-31.1"
 /clone_id="RP5-836J3"
 /clone_1lb="RPL1-5"
 1160..1786
 /note="match: GSS: Em: B16380"
 complement(2068..2500)
 /note="match: GSS: Em: A089628"
 2554..3072
 /note="match: GSS: Em: A0118130"
 complement(3824..4133)
 /note="match: GSS: Em: A089896"
 4017..4402
 /note="match: GSS: Em: A063212"
 4017..4312
 /note="match: GSS: Em: A0059070"
 4023..4278
 /note="match: GSS: Em: A0063209"
 6000..6045
 /note="2 copies 23 mer 100% conserved"
 6637..6906
 /note="ORSL repeat: matches 190..467 of consensus"
 9859..10375
 /note="MLTIF repeat: matches 3..536 of consensus"
 13112..13891
 /note="LTP17 repeat: matches 1..780 of consensus"
 15130..15173
 /note="22 copies 2 mer tg 93% conserved"
 15133..15172
 /note="4 copies 10 mer gtgtgtgtgt 95% conserved"
 23083..23118
 /note="18 copies 2 mer tg 80% conserved"
 23328..23713
 /note="WSTD repeat: matches 1..394 of consensus"
 26442..26509
 /note="134 copies 2 mer gt 66% conserved"
 complement(26847..27324)
 /note="match: GSS: Em: A0835456"
 27344..27699
 /note="match: GSS: Em: A0479571"
 31575..32054
 /note="match: STS: Em: AL009625"
 32094..32491
 /note="MLT2D repeat: matches 1..490 of consensus"
 32492..32857
 /note="match: GSS: Em: A0140889"
 32521..32580
 /note="6 copies 10 mer gatagataga 80% conserved"
 32528..32587
 /note="15 copies 4 mer atag 96% conserved"
 32591..32639
 /note="MLT2D repeat: matches 505..553 of consensus"
 38429..38876
 /note="match: GSS: Em: A0146760"
 39727..39756
 /note="6 copies 5 mer tgttt 96% conserved"
 40414..40453
 /note="20 copies 2 mer ca 97% conserved"
 42939..42989
 repeat_region

```
repeat_region /note="MER91A repeat: matches 95. .152 of consensus"
misc_feature /note="2 copies 48 mer 85% conserved"
misc_feature complement(43694. .44158)
misc_feature /note="match: GSS: Em:AQ437799"
repeat_region 44168. .44654
repeat_region /note="match: GSS: Em:AQ564589"
misc_feature 45309. .45340
repeat_region /note="16 copies 2 mer tt 93% conserved"
misc_feature 45341. .45660
repeat_region /note="match: STS: Em:G24052"
repeat_region 45473. .45610
repeat_region /note="6 copies 23 mer 67% conserved"
repeat_region 46017. .46066
repeat_region /note="25 copies 2 mer ac 78% conserved"
repeat_region 46019. .46066
repeat_region /note="12 copies 4 mer acac 79% conserved"
repeat_region 49244. .49451
repeat_region /note="MER58A repeat: matches 1. .223 of consensus"
repeat_region 51847. .51898
repeat_region /note="13 copies 4 mer cctc 73% conserved"
misc_feature 51849. .51898
repeat_region /note="25 copies 2 mer ct 74% conserved"
misc_feature complement(52909. .53347)
repeat_region /note="match: GSS: Em:AQ440464"
repeat_region 56014. .56544
repeat_region /note="MER68B repeat: matches 1. .568 of consensus"
repeat_region 57696. .57814
repeat_region /note="MER69 repeat: matches 2. .129 of consensus"
repeat_region 58171. .58704
repeat_region /note="MLT1E repeat: matches 3. .548 of consensus"
misc_feature 66786. .67150
repeat_region /note="match: GSS: Em:AQ053163"
repeat_region 67192. .67597
repeat_region /note="MER7A repeat: matches 1. .346 of consensus"
misc_feature complement(68816. .69390)
repeat_region /note="match: GSS: Em:AQ549150"
repeat_region 69397. .69432
repeat_region /note="9 copies 4 mer taag 86% conserved"
misc_feature 69519. .70015
repeat_region /note="match: GSS: Em:AQ045862"
repeat_region 69911. .69986
repeat_region /note="19 copies 4 mer taat 65% conserved"
repeat_region 70512. .71004
repeat_region /note="MER4A repeat: matches 110. .656 of consensus"
repeat_region 71303. .71412
repeat_region /note="MER4A repeat: matches 1. .110 of consensus"
misc_feature complement(73648. .74145)
repeat_region /note="match: GSS: Em:AQ458708"
repeat_region 75280. .75333
repeat_region /note="27 copies 2 mer tt 72% conserved"
repeat_region 76040. .76115
repeat_region /note="19 copies 4 mer aaga 97% conserved"
repeat_region 76046. .76115
repeat_region /note="35 copies 2 mer ga 75% conserved"
repeat_region 76047. .76126
repeat_region /note="8 copies 10 mer aagaagaaga 76% conserved"
repeat_region 78923. .79419
repeat_region /note="MLT1F repeat: matches 67. .541 of consensus"
repeat_region 80005. .80032
repeat_region /note="7 copies 4 mer atac 96% conserved"
repeat_region 82920. .83118
repeat_region /note="MER91A repeat: matches 2. .195 of consensus"
repeat_region 83123. .83171
repeat_region /note="MER91A repeat: matches 138. .186 of consensus"
repeat_region 83659. .84165
repeat_region /note="LIPB2 repeat: matches 5634. .6155 of consensus"
repeat_region 87050. .87201
repeat_region /note="MER4A repeat: matches 1. .148 of consensus"
repeat_region 87197. .87536
repeat_region /note="MER4A repeat: matches 394. .728 of consensus"
misc_feature 91670. .92092
/note="match: GSS: Em:AQ275888"
```

```
misc_feature 91670. .92091
misc_feature /note="match: GSS: Em:AQ279665"
misc_feature 91670. .92003
misc_feature /note="match: GSS: Em:AQ278706"
repeat_region 93141. .94747
repeat_region /note="MSTA-internal repeat: matches 1. .1637 of consensus"
repeat_region 94761. .95183
repeat_region /note="MSTB repeat: matches 1. .426 of consensus"
repeat_region 95891. .96457
repeat_region /note="LIMC1 repeat: matches 5219. .5796 of consensus"
repeat_region 96486. .96839
repeat_region /note="MER74A repeat: matches 165. .524 of consensus"
repeat_region 96940. .97505
repeat_region /note="LIMC1 repeat: matches 5791. .6325 of consensus"
repeat_region 100936. .100997
repeat_region /note="LIMC1 repeat: matches 230. .291 of consensus"
repeat_region 105831. .106199
repeat_region /note="LOR1B repeat: matches 1. .461 of consensus"
repeat_region 106592. .107055
repeat_region /note="LIMB5 repeat: matches 5688. .6174 of consensus"
repeat_region 107066. .107248

Query Match 72.0%; Score 359.8; DB 9; Length 143590;
Best Local Similarity 88.3%; Pred. No. 4.8e-108;
Matches 414; Conservative 0; Mismatches 52; Indels 3; Gaps 2;

QY 34 AAAAACTCGAGATACACCAATATCTTGCAACATGAGACAGACGACTAGCTGATTTCC 93
DB 13925 AAAAAATAGACAGCTACAGATATGAGACATCTAGAGATAGAGCTAGCTGATTTCC 13866
QY 94 TAGCCGAGCTAAGAAATCCCTTAAGCTAGCTGGGAAGTACACGCTCACTTTAAACAC 153
DB 13865 TAGCCGAGCTAAGAAATCCCTTAAGCTAGCTGGGAAGTACACGCTCACTTTAAACAT 13806
QY 154 GGGGCTTGCACTTACCTACACCTGACCAATCAATGAGAGCTCACTAAATGCTAATTAGG 213
DB 13805 GGGGCTTGCACTTACCTACACCTGACCAATCAATGAGAGCTCACTAAATGCTAATTAGG 13746
QY 214 CAAGAAGAGAGTAAAGAAATAGCAATCATCTATTTGCTGAGACAGACAGAGGGA 273
DB 13745 CAAGAAGAGAGTAAAGAAATAGCAATCATCTATTTGCTGAGACAGAGGGA 13686
QY 274 CAACAATCGGAGATATTAACCCAGGACTGAGCTGCAACAGAGAGCCCTTTGGGCTCC 333
DB 13685 CAAGAATCGGAGATATTAACCCAGGACTGAGCTGCAACAGAGAGCCCTTTGGGCTCC 13627
QY 334 CTTCCTTTGTATGAGAG--CTGTTTCAATGCTATTCTCTATTAAATCTTCAACTG 391
DB 13626 CTTCCTTTGTATGAGAGCTGTTTCACTCTATTCTCTATTAAATCTTCAACTG 13567
QY 392 CACTCTTGTGCTGCTGCTTTCTTAAGCTGAGCTGAGCTTTTGTCTCACCGTCCACTACT 451
DB 13566 CACTCTTGTGCTGCTGCTTTCTTAAGCTGAGCTGAGCTTTTGTCTCACCGTCCACTACT 13507
QY 452 GCTGTTGGCACACCGCAGACCTGCGGCTGACTCCCATCCCTCTGAT 500
DB 13506 GCTGTTGGCACACCGTCCAGACCGCTGACTCTTCACTCCCTCTGAT 13458

RESULT 10
AC103595/c 119406 bp DNA linear HTG 29-NOV-2001
LOCUS Homo sapiens chromosome 1 clone RP11-415P10, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.
ACCESSION AC103595
VERSION AC103595.1 GI:17149460
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULTOP; HTGS_ACTIVEPIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 119406)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
```


JOURNAL	Direct Submission
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 119406)
TITLE	Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
JOURNAL	Submitted (29-NOV-2001) Genome Center, University of Washington
COMMENT	Box 352145, Seattle, WA 98195, USA.
	----- Genome Center -----
	Center: University of Washington Genome Center
	Center Code: UWGC
	Web site: http://www.genome.washington.edu
	Contact: uwgchgs@u.washington.edu
	----- Project Information -----
	Center project name: chr-1
	Center clone name: RP11-415P10 (ec0362)
	----- Summary Statistics -----
	Sequencing vector: plasmid; L03752; 100% of reads
	Chemistry: Dye-terminator ET; 32% of reads
	Chemistry: Dye-terminator Big Dye; 8% of reads
	Assembly program: Phrap; version 0.990319
	Consensus quality: 109646 bases at least Q40
	Consensus quality: 115473 bases at least Q30
	Consensus quality: 117774 bases at least Q20
	Insert size: 194800; 40.1% error; agarose-fp
	Insert size: 118506; sum-of-contigs
	Quality coverage: 5.7x in Q20 bases; agarose-fp
	Quality coverage: 9.3x in Q20 bases; sum-of-contigs

	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 10 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.

	* 1 4544: contig of 4564 bp in length
	* 4565 4664: gap of unknown length
	* 4665 12930: contig of 8266 bp in length
	* 12931 13030: gap of unknown length
	* 13031 20784: contig of 7754 bp in length
	* 20785 20884: gap of unknown length
	* 20885 29093: contig of 8209 bp in length
	* 29094 29193: gap of unknown length
	* 29194 40511: contig of 1116 bp in length
	* 40512 40611: gap of unknown length
	* 40612 52409: contig of 11796 bp in length
	* 52410 52509: gap of unknown length
	* 52510 64875: contig of 12266 bp in length
	* 64876 64876: gap of unknown length
	* 64876 83398: contig of 18523 bp in length
	* 83399 83498: gap of unknown length
	* 83499 100056: contig of 16558 bp in length
	* 100057 100156: gap of unknown length
	* 100157 119406: contig of 19250 bp in length.
FEATURES	Location/Qualifiers
source	1. .119406
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="1"
	/clone="RP11-415P10"
	/clone_lib="RP11 human BAC library 11"
	1. .4564
misc_feature	/note="assembly_name:Contig:9"
misc_feature	4665. .12930
	/note="assembly_name:Contig30"
misc_feature	13031. .20784
	/note="assembly_name:Contig31"
misc_feature	20885. .29093
	/note="assembly_name:Contig32"
misc_feature	29194. .40511
	/note="assembly_name:Contig33"
misc_feature	40612. .52409

misc_feature	/note="assembly_name: Contig34"	52510..64775
misc_feature	/note="assembly_name: Contig35"	64876..83398
misc_feature	/note="assembly_name: Contig36"	83499..100056
misc_feature	/note="assembly_name: Contig37"	100157..119406
misc_feature	/note="assembly_name: Contig38"	
BASE COUNT	35588 a 23913 c 23787 g 35137 t	961 others
ORIGIN		
Query Match	71.6%; Score 358.2; DB 2; Length 119406;	
Best Local Similarity	88.1%; Pred. No. 1.6e-107;	
Matches 413; Conservative	0; Mismatches 53; Indels 3; Gaps 2;	
Qy	34	AAACGCTTGAGATATACGACATTTATCTTGCACTGAGAGACAGACTGAGTGAATTTCC 93
Db	37607	AAATATATAGAGCTATACAGATTTAGAGACATCTAGAGATAGATGAGACTGAGTGAATTTCC 37548
Qy	94	TAGCGCACTTAAAGATTCCTTAAGCTTACCTTGGAAGGTGACACAGTCCACTTTAAACAC 153
Db	37547	TAGCGCACTTAAATATCCCTTAAGCTTACCTTGGAAGGTGACACAGTCCACTTTAAACAT 37488
Qy	154	GGGGCTTGCAACTTAAAGCTCAACCTGACCAATCAGAGAGCTCAGTAAATCTTAATTAG 213
Db	37487	GGGGCTTGCAACTTAAAGCTCAACCTGACCAATCAGAGAGCTCAGTAAATCTTAATTAG 37428
Qy	214	CAAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTTGAAGACACAGAGAGGGA 273
Db	37427	CAAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTTGAAGACACAGAGAGGGA 37368
Qy	274	CAACATTCGGGATTTAAACCCAGGCACTTCAGCTGCGAACAGAGCCGCCCTTTGGGTC 333
Db	37367	CAAGATTCGGGATTTAAACCCAGGCACTTCAGCTGCGAACAGAGCCGCCCTTTGGGTC 37309
Qy	334	CTTCCCTTTGATGAGGAG--CTGTTTCATCTATTTCACTCTATTAATCTTGAACATG 391
Db	37308	CTTCCCTTTGATGAGGAGCTGTTTCATCTCTATTTCACTCTATTAATCTTGAACATG 37249
Qy	392	CACCTCTCTGCTGCATGTTCTTAAGGCTGAGCTGAGCTTTTGCTCAACGTCACACT 451
Db	37248	CACCTCTCTGCTGCATGTTCTTAAGGCTGAGCTGAGCTTTTGCTCAACGTCACACT 37189
Qy	452	GCTGTTTCCACACCGCAGACCTGCGCTGATCCCTCCCTCGGAT 500
Db	37188	GCTGTTTCCACACCGCTCCAGACCGCTGATCCCTCCCTCGGAT 37140
RESULT 11		
AL353584_0/c		
WPCOMMENT		
Sequence split into 6 fragments	LOCUS AL353584	Accession AL353584
Fragment Name	Begin	End
AL353584_0	1	110000
AL353584_1	100001	210000
AL353584_2	200001	310000
AL353584_3	300001	410000
AL353584_4	400001	510000
AL353584_5	500001	536214
LOCUS	AL353584	536214 bp DNA linear HTG 19-JUL-2003
DEFINITION	Homo sapiens chromosome X clone RP13-185F3, *** SEQUENCING IN	
ACCESSION	AL353584	
VERSION	AL353584.11	GI:15020381
KEYWORDS	HTG; HTGS_PHASE2; HTGS_CANCELLED.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Enxaiyotia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
TITLE	1 (bases 1 to 536214)	
	Wilson, S.	
	Direct Submission	

JOURNAL

Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:13121369.

COMMENT

----- Genome Center
Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: bH185F3

----- Summary Statistics

Assembly program: XGAP4, version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator ET-amersham; 28% of reads

Dye-terminator Big Dye; 71% of reads

Consensus quality: 167016 bases at least Q40

Consensus quality: 168523 bases at least Q30

Insert size: 536214; sum-of-contigs

Insert size: 160797; 5.5% error; agarose-fp

Quality coverage: 2.13x in Q20 bases; sum-of-contigs Quality

coverage: 7.40x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

Location/Qualifiers

1..356214

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="RP13-185F3"

/clone_1ib="RPCT-13.1"

misc_feature

1..356214

/note="assembly_fragment:03802"

clone_end:SP6

vector_side:left

clone_end:T7

vector_side:left"

BASE COUNT 169781 a 102600 c 103981 g 159850 t 2 others

ORIGIN

Query Match 71.6%; Score 357.8; DB 2; Length 110000;

Best Local Similarity 92.0%; Pred. No. 2.2e-107;

Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACAGACTAGCTGATTTCTAGGCGGACTAAGATCCTTAAGCTAGCTGGGA 127

Db 38246 TGAGAGACAGACTAGCTGATTTCTAGGCGGACTAAGATCCTTAAGCTAGCTGGGA 38187

QY 128 AGGTGACACAGCTCACTTTAAACACGGGCTTGCACTTAGCTCAGACCTGACCAATCA 187

Db 38186 AGGTGACACAGCTCACTTTAAACACGGGCTTGCACTTAGCTCAGACCTGACCAATCA 38127

QY 188 GAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCAATCATCT 247

Db 38126 GAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCAATCATCT 38067

QY 248 ATTGCTGAGAGACAGACAGAGGACCAACATCGGATATTAACCCAGGACTTCGAGCT 307

Db 38066 ATTGCTGAGAGACAGACAGAGGACCAACATCGGATATTAACCCAGGACTTCGAGCT 38007

QY 308 GGCACACAGACGCCCCCTTTGGGCTCTTCCCTTTGATGGAG--CTGTTTCATGCTGA 365

Db 38006 GGCACACAGACGCCCCCTTTGGGCTCTTCCCTTTGATGGAGCTCTGTTTCATGCTGTG 37948

QY 366 TTTCACCTCTAATAATCTTGCACTGACCTCTTGCTGATGTTTCTTAAGGCTGAGC 425

Db 37947 TTTCACCTCTAATAATCTTGCACTGACCTCTTGCTGATGTTTCTTAAGGCTGAGC 37888

QY 426 TGAGCTTTTGTCTGACGCTGACCACTGCTGTTTGCACACCGGAGACTGCGCTGACT 485

Db 37887 TGAGCTTTTGTCTGACGCTGACCACTGCTGTTTGCACCGGAGACTGCGCTGACT 37828

QY 486 CCATCCCTCTGAGT 500

Db 37827 TCCATCCTTCAGAT 37813

RESULT 12

CNS06C7R/c

LOCUS

DEFINITION

CNS06C7R 169029 bp DNA linear PRI 11-MAY-2001

Human chromosome 14 DNA sequence BAC R-254B15 of library RPCT-11

from chromosome 14 of Homo sapiens (human), complete sequence.

ACCESSION

VERSION

AL330800.4 GI:14041153

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : C-258117
Downstream BAC (overlapping the SP6 end) : R-779M2 (AC=AL451071)

Assembly Statistics

Assembly program: Phrap; version 2.0

Quality coverage: 9.37x in Q20 bases; sum-of-contigs

Summary Statistics

Overall quality chart :

Range : bases

0 - 9

10 - 19

20 - 29

30 - 39

40 - 49

50 - 59

60 - 69

70 - 79

80 - 89

90 - 99

86603

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

source

Location/Qualifiers

1..169029

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="14"

/clone="R-254B15"

/clone_1ib="RPCT-11"

BASE COUNT

54704 a 28998 c 29595 g 55732 t

ORIGIN

Query Match 71.6%; Score 357.8; DB 9; Length 169029;
 Best Local Similarity 92.0%; Pred. No. 2 3e-107;
 Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACAGAGCTAGCTGATTTCTTGAAGCCGACATGAAGTTCCTTAAGCTTACTGCTGGGA 127
 Db 22071 TGAGAGACAGAGCTAGCTGATTTCTTGAAGCCGACATGAAGTTCCTTAAGCTTACTGCTGGGA 22012

QY 128 AGGAGACAGCTGACCTTTAAACAGCGGGCTTGCACTTGTAGCTCACACCTGACCAATCA 187
 Db 22011 AGGAGACAGCTGACCTTTAAACAGCGGGCTTGCACTTGTAGCTCACACCTGACCAATCA 21952

QY 188 GAGAGCTCACTAAATGCTAATTAAGCAGAGAGTAAAGAAATAGCCATCATCT 247
 Db 21951 GAGAGCTCACTAAATGCTAATTAAGCAGAGAGTAAAGAAATAGCCATCATCT 21892

QY 248 ATTGCTGAGAGCAGAGAGGAGGACACATGGGATATTAACCCAGGATTCGAGCT 307
 Db 21891 ATTGCTGAGAGCAGAGGAGGAGGACACATGGGATATTAACCCAGGATTCGAGCT 21832

QY 308 GGGACAGCAGACCCCTTTGGGTCCCTTCCCTTGGATGGAG--CTGTTTCATGCTA 365
 Db 21831 GGGACAGCAGACCCCTTTGGGTCCCTTCCCTTGGATGGAGCTGTTTCATGCTA 21773

QY 366 TTTCACCTATTAATCTTGAACCTGACTCTTGTGCTTCATGTTTCTTACGGCTCGAGC 425
 Db 21772 TTTCACCTATTAATCTTGAACCTGACTCTTGTGCTTCATGTTTCTTACGGCTCGAGC 21713

QY 426 TGAAGTTTGTCTCACTGCTGCAACCTGCTGTTTGGCCACACCGGAGACTGCGCTGACT 485
 Db 21712 TGAAGTTTGTCTGCTGCTGCAACCTGCTGTTTGGCCACATTAACGACCGCGCTGACT 21653

QY 486 CCCATCCCTCGGAT 500
 Db 21652 CCCATCCCTCGGAT 21638

RESULT 13
 AL392173/c 170746 bp JMA linear PRI 05-APR-2001
 LOCUS Human DNA sequence from clone RP11-158M9 on chromosome Xq26.1-27.1
 DEFINITION Contains STS and GSSs, complete sequence.
 ACCESSION AL392173
 VERSION AL392173.8 GI:11064224
 KEYWORDS HMG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 170746)
 Bray-Allen, S.
 Direct Submission
 Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humgery@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Oct 31, 2000 this sequence version replaced gi:10800700.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone conigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/ChrX
 This sequence is the entire insert of clone RP11-158M9 This
 sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-158M9 is from
 the library RPEC11.1 constructed by the group of Pieter de Jong.
 For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBAC3.6.

FEATURES

source

Location/Qualifiers

1..170746
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q26.1-27.1"
 /clone="RP11-158M9"
 /clone_1fb="RPCT-11.1"
 688..1038
 /note="match: GSS: Em:AQ314307"
 2318..2343
 /note="13 copies 2 mer ca 100% conserved"
 7251..7937
 /note="L1PA6 repeat: matches 4764..5447 of consensus"
 7938..8568
 /note="L1PA6 repeat: matches 5507..6139 of consensus"
 9828..10340
 /note="match: GSS: Em:AZ368366"
 14268..14619
 /note="MER34 repeat: matches 4..301 of consensus"
 20566..21107
 /note="MLR2B repeat: matches 1..448 of consensus"
 27734..27963
 /note="match: STS: Em:HS210VD10"
 27963..27998
 /note="18 copies 2 mer ac 100% conserved"
 27967..28062
 /note="match: STS: Em:HS210VD10"
 29226..29569
 /note="THE1C repeat: matches 1..371 of consensus"
 29570..30022
 /note="THE1C internal repeat: matches 1..445 of consensus"
 30335..31434
 /note="THE1C-internal repeat: matches 445..1580 of
 consensus"
 31435..31807
 /note="THE1C repeat: matches 1..371 of consensus"
 35205..35669
 /note="match: GSS: Em:AQ493428"
 37464..38246
 /note="L1R17 repeat: matches 1..780 of consensus"
 39054..39117
 /note="132 copies 2 mer gt 92% conserved"
 41432..41546
 /note="L1R29 repeat: matches 516..619 of consensus"
 41877..42007
 /note="L1R29 repeat: matches 3..149 of consensus"
 51113..51140
 /note="14 copies 2 mer tc 89% conserved"
 51268..51318
 /note="H15 repeat: matches 30..76 of consensus"
 54978..55013
 /note="18 copies 2 mer ca 91% conserved"
 58352..59188
 /note="THE1B-INTERNAL repeat: matches 1..842 of consensus"
 61106..61272
 /note="WS1B-internal repeat: matches 15..186 of consensus"
 61578..61999
 /note="WS1B-internal repeat: matches 186..585 of
 consensus"
 62003..62098
 /note="WS1-INTERNAL repeat: matches 594..689 of consensus"
 63045..63698

repeat_region /note="HERV16 repeat: matches 4. .664 of consensus"
63927. .64774 /note="HERV16 repeat: matches 716. .1577 of consensus"
repeat_region 64789. .64832 /note="22 copies 2 mer aa 81% conserved"
repeat_region 64916. .65612 /note="HERV16 repeat: matches 1817. .2518 of consensus"
repeat_region 65613. .66886 /note="THER1-INTERNAL repeat: matches 124. .1401 of consensus"
misc_feature 66475. .66484 /note="1372 bp of IS186 transposon (X03123) removed here. This sequence represents the duplicated flanking sequence of the IS186."
misc_feature 70601. .70772 /note="match: GSS: Em:AQ077330"
misc_feature 70734. .71229 /note="match: GSS: Em:AQ628837"
misc_feature 80000. .80512 /note="match: GSS: Em:AQ284023
match: STS: Em:G61704"
misc_feature complement(81135. .81789)
/note="match: GSS: Em:AQ240981"
81193. .81655 /note="match: GSS: Em:AQ704174"
83625. .89762 /note="LIPAS repeat: matches 11. .6145 of consensus"
91802. .91895 /note="LIPAS repeat: matches 6053. .6146 of consensus"
94922. .101052 /note="LIPAS repeat: matches 11. .6145 of consensus"
102779. .102828 /note="LIPAS repeat: matches 11. .6145 of consensus"
102779. .102828 /note="25 copies 2 mer ac 82% conserved"
misc_feature complement(103577. .104157)
/note="match: GSS: Em:AQ775154"
105177. .105697 /note="match: GSS: Em:B63384"
105842. .106341 /note="match: GSS: Em:AQ522879"
107797. .108270 /note="match: GSS: Em:AQ002183"
116826. .116918 /note="MER39 repeat: matches 13. .521 of consensus"
116826. .116918 /note="MER39 repeat: matches 1. .94 of consensus"
complement(122839. .123304)
/note="match: GSS: Em:AQ002183"
126653. .126696 /note="22 copies 2 mer ac 93% conserved"
127908. .127950 /note="LIPAS repeat: matches 6103. .6145 of consensus"
complement(129916. .130692)
/note="match: GSS: Em:AQ288149"
complement(130168. .130745)
/note="match: GSS: Em:AQ293756"
complement(131672. .132168)
/note="match: GSS: Em:AQ466443"
131821. .131846 /note="13 copies 2 mer ta 96% conserved"
132163. .132600 /note="match: GSS: Em:AQ829322"
133493. .133664 /note="LIPAS repeat: matches 30. .202 of consensus"
136095. .136245 /note="LIPAS repeat: matches 6142. .6289 of consensus"
complement(140979. .141356)
/note="match: STS: Em:HSC28H08"
146003. .146304 /note="LIPAS repeat: matches 1. .310 of consensus"
146351. .146421 /note="LIPAS repeat: matches 448. .516 of consensus"
157021. .157123 /note="MER34 repeat: matches 76. .173 of consensus"
159023. .159689 /note="HERV16 repeat: matches 562. .1217 of consensus"

repeat_region 15978. .160033 /note="HERV16 repeat: matches 1217. .1272 of consensus"
repeat_region 164301. .166134 /note="LIPAS repeat: matches 1841. .3688 of consensus"
repeat_region 166458. .166509 /note="26 copies 2 mer ga 73% conserved"
repeat_region 166568. .169037 /note="LIPAS repeat: matches 3681. .6148 of consensus"
misc_feature 169389. .169794 /note="match: GSS: Em:AQ092454"
misc_feature complement(170139. .170746)
/note="match: GSS: Em:AQ389950"
misc_feature complement(170286. .170416)
/note="match: GSS: Em:AQ385692"
BASE COUNT 53940 a 32801 c 34298 g 49707 t
ORIGIN

Query Match 71.6%, Score 357.8, DB 9, Length 170746;
Best Local Similarity 92.0%; Pred.No. 2.3e-107;
Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACGAGTCTAGTCTGATTTCTAGGCGGACGCTAAGATCCCTAAGCTAGCTGGGA 127
Db 38246 TGAGAGACGAGTCTAGTCTGATTTCTAGGCGGACGCTAAGATCCGAGCTAGCTGGGA 38187
QY 128 AGGTGACACAGTCCACCTTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAATCA 187
Db 38186 AGGTGACGATCATCCTTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAATCA 38127
QY 188 GAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGTAAAGAAATAGCAATCATCT 247
Db 38126 GAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGTAAAGAAATAGCAATCATCT 38067
QY 248 ATTGCGTGAAGAGACAGGAGGAGCAACATCGGATTAATCCAGGCAATTCAGCT 307
Db 38066 ATTGCGTGAAGAGACAGGAGGAGCAACATCGGATTAATCCAGGCAATTCAGCT 38007
QY 308 GGGACAGCAGCCCCCTTTGGGCTCCCTTCTTGTATGGAG--CTGTTTCATGCTA 365
Db 38006 GGGACAGGCA--ACCCCTTTGGGCTCCCTTCTTGTATGGAGCTGTTTCATGCTG 37948
QY 366 TTTCATCTAATTAATCTTGCAATCGACTCTTGTGCTCATGTTCTTAAGCTGAGC 425
Db 37947 TTTCATCTAATTAATCTTGCAATCGACTCTTGTGCTCATGTTCTTAAGCTGAGC 37888
QY 426 TGAGCTTTTGTCAACGCTCAACGCTGCTTTGGCAACGCGGACGCTGACT 485
Db 37887 TGAGCTTTTGTCAACGCTCAACGCTGCTTTGGCCGCGTGGAGACCGCGTACT 37828
QY 486 CCCATCCCTCTGAT 500
Db 37827 TCCATCCCTCCAGT 37813

RESULT 14
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS AC093531 Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531.2 GI:16945981
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
RECORD 2 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint

REFERENCE	GENOME INSTITUTE, 2800 MITCHELL DRIVE, WALNUT CREEK, CA 94598, USA
AUTHORS	3 (bases 1 to 163803)
JOURNAL	DOE JOINT GENOME INSTITUTE AND STANFORD HUMAN GENOME CENTER.
COMMENT	DIRECT SUBMISSION Submitted (16-NOV-2001) DOE JOINT GENOME INSTITUTE, 2800 MITCHELL DRIVE, WALNUT CREEK, CA 94598, USA On Nov 16, 2001 this sequence version replaced gi:15183820. Draft Sequence Produced by DOE JOINT GENOME INSTITUTE www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.sngc.stanford.edu Quality: Phrap Quality >=40 99.8% of Sequence; Estimated Total Number of Errors :s 0.3.
FEATURES	Location/Qualifiers
SOURCE	1. 163803 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="5" /clone="RP11-405L7"
BASE COUNT	54797 a 31372 c 29086 g 48548 t
ORIGIN	
Query Match	71.5%; Score 357.4; DB 9; Length 163803; Best Local Similarity 91.2%; Pred. No. 3,1e-107; Matches 402; Conservative 0; Mismatches 36; Indels 3; Gaps 2;
Db 120271	TGCAACTGAGACAGACAGACTAGCTGAGATTCCTTAGGTCGACTAAGATCCCTAAGCCTTAG 121
Qy 122	CTGGGAAGGTGACACGCTCCACCTTTAAACAGGGGGCTTGCAACTTAGCTCACCTGAC 181
Db 120211	CTGGGAAGGTGACCTCATCCACTACCTTTAAACATGGGGCTTGCAACTTAGCTCACCTGAC 120152
Qy 182	CAATCAGAGAGCTCACTAAATGCTAATTAGGCMAAGACAGAGGTAAAGAAATAGCCAA 241
Db 120151	CAATCAGAGAGCTCACTAAATGCTAATTAGGCMAAAGAGAGGCAAAAGAAATAGCCAA 120092
Qy 242	TCATCTAATTGCTTGAGAGACAGACAGAGGAGCAACATCGGAGATTAATACCCAGGCATT 301
Db 120091	TCATCTAATTGCTTGAGAGACAGACAGGAGGACCAAGATCGGAGATTAATACCCAGGCATT 120032
Qy 302	CGAGCTGGCAACAGACAGCCCCCTTTGGGTGCCCTCCCTTGATATGGAGAG--CTGTTTC 359
Db 120031	TCAGCTGGCAAGGCA-ACCCCTTTGGGTGCCCTCCCTCTCTCTTTGATATGGAGCTCTGTTTC 119973
Qy 360	ATGCTAATTCACCTATTAATCTTGCAACTGCACCTCTCTGATCCATGTTCTTAACGGC 419
Db 119972	ACTCTAATTCACCTATTAATCTTGCAACTGCACCTCTCTGATCCGCTGTTTGTTAAGGC 119913
Qy 420	TGGAGCTGAGCTTTTGCTCAGCCGTCACCACTGCTGTTTGCCACCAACCGCAGACCTGCCG 479
Db 119912	TCAGCTGAGCTTTTGCTGCTGCTGCTCACCACCTGCTGTTTGCCACCAATCGAGACCACTG 119853
Qy 480	CTGACTCCCATCCCTCTGGAT 500
Db 119852	CTGACTTCATCCCTCCAGAT 119832
RESULT 15	
AC010888	
LOCUS	AC010888 191863 bp DNA linear PRI 07-NOV-2001
DEFINITION	Homo sapiens BAC clone RP11-418N16 from 2, complete sequence.
ACCESSION	AC010888
VERSION	AC010888.12 GI:13992760
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
AUTHORS	1 (bases 1 to 191863) Sulston, J.E. and Waterston, R.
TITLE	Toward a complete human genome sequence

JOURNAL-MEDLINE	Genome Res. 8 (11), 1097-1108 (1998)
REFERENCE	99063792
AUTHORS	PUBMED
TITLE	9847074
JOURNAL	2 (bases 1 to 191863)
REFERENCE	Margolis, C., Abdoct, A. and Sawyer, C.
AUTHORS	The sequence of Homo sapiens BAC clone RP11-418M16
TITLE	Unpublished
JOURNAL	3 (bases 1 to 191863)
REFERENCE	Waterston, R.H.
AUTHORS	Direct Submission
TITLE	Submitted (25-SEP-1999) Genome Sequencing Center, Washington
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 191863)
AUTHORS	Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (08-MAY-2001) Genome Sequencing Center, Washington
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
AUTHORS	5 (bases 1 to 191863)
TITLE	Waterston, R.
JOURNAL	Direct Submission
REFERENCE	Submitted (09-AUG-2001) Department of Genetics, Washington
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE	6 (bases 1 to 191863)
JOURNAL	Waterston, R.
REFERENCE	Direct Submission
AUTHORS	Submitted (07-NOV-2001) Department of Genetics, Washington
TITLE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL	On May 8, 2001 this sequence version replaced gi113431186.
COMMENT	

Center: Washington University Genome Sequencing Center
 Center code: WMSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapieris@wustl.wustl.edu
 Summary Statistics
 Center project name: H_NH0418N16

NOTICE: This sequence may not represent the entire insect of choice. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION: Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/9sc>

SOURCE INFORMATION: The RPc1-11 human BAC library was made from the blood of one male donor, as described by Ogoeawaka, K., Moon, P. Y., Zhao, B., Fritgen, E., Tatenno, M., Caranese, J. J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is RP11-434M17, 200 bp overlap. Actual start of this clone is at base position 135 of RP11-418N16; actual end is at base position 191863 of RP11-418N16.

The sequence RP11-418N16 between bases 117880 to 117930 is covered by a single plasmid subclone that has some low quality regions.

FEATURES

source

Location/Qualifiers

```
1. 191863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-418N16"
/clone_1lb="RPC1-11"
repeat_region
387..700
/rpc_family="Alu"
repeat_region
906..1517
/rpc_family="MaLR"
repeat_region
2319..2372
/rpc_family="AT_rich"
repeat_region
2456..3313
/rpc_family="L1"
repeat_region
3314..3345
/rpc_family="(TAA)n"
repeat_region
3346..4533
/rpc_family="L1"
repeat_region
4560..4679
/rpc_family="CT-rich"
repeat_region
4680..6289
/rpc_family="SVA"
misc_feature
4839..5815
/note="CG island (GC=68.7, o/e=0.70, #CGs=68)"
repeat_region
6221..6330
/rpc_family="L1"
repeat_region
6331..6642
/rpc_family="Alu"
repeat_region
6643..7867
/rpc_family="L1"
repeat_region
7870..8126
/rpc_family="Mariner"
repeat_region
8127..8491
/rpc_family="L1"
repeat_region
8493..9037
/rpc_family="ERVU"
repeat_region
9038..10102
/rpc_family="ERVU"
repeat_region
10097..10959
/rpc_family="ERVU"
repeat_region
10960..11329
/rpc_family="MaLR"
repeat_region
12986..13361
/rpc_family="MaLR"
repeat_region
13362..15300
/rpc_family="ERVU"
repeat_region
15300..15696
/rpc_family="ERVU"
repeat_region
15720..16137
/rpc_family="Tc2"
repeat_region
16149..16206
/rpc_family="Mariner"
repeat_region
16210..16313
/rpc_family="Tc2"
repeat_region
16571..16706
/rpc_family="L2"
repeat_region
17680..17752
/rpc_family="L1"
repeat_region
17777..17859
/rpc_family="MIR"
repeat_region
18083..18285
/rpc_family="L1"
repeat_region
18546..18879
/rpc_family="L1"
repeat_region
22173..22474
/rpc_family="Alu"
repeat_region
23231..24056
/rpc_family="MER2_type"
repeat_region
24453..24476
```

```
/rpc_family="AT_rich"
repeat_region
24477..24768
/rpc_family="L1"
repeat_region
24769..24801
/rpc_family="AT_rich"
repeat_region
24908..25040
/rpc_family="MIR"
repeat_region
25445..25761
/rpc_family="MER1_type?"
repeat_region
26238..26534
/rpc_family="Alu"
repeat_region
26605..26932
/rpc_family="MER1_type?"
repeat_region
27139..27264
/rpc_family="MIR"
misc_feature
27364..27706
/note="similar to EST AW892638 (NTD:8056843)"
repeat_region
27715..27889
/rpc_family="MIR"
repeat_region
28246..28385
/rpc_family="L2"
repeat_region
28665..28689
/rpc_family="AT_rich"
repeat_region
28925..29245
/rpc_family="Alu"
repeat_region
29844..30032
/rpc_family="L1"
repeat_region
30033..30079
/rpc_family="US"
repeat_region
30080..30487
/rpc_family="L1"
repeat_region
30488..30650
/rpc_family="MER1_type"
repeat_region
30651..30929
/rpc_family="Alu"
repeat_region
30930..30974
/rpc_family="MER1_type"
repeat_region
30975..31085
/rpc_family="L1"
repeat_region
31211..31417
/rpc_family="MIR"
repeat_region
31455..31632
```

Query Match 71.5%; Score 357.4; DB 9; Length 191863;
Best Local Similarity 91.2%; Pred No. 3.1e-107;
Matches 402; Conservative 0; Mismatches 36; Indels 3; Gaps 2;

```
QY 62 TGCACCTGAGAGACAGGACTGAGTTCCTTAGGCCGACTAAGATCCCTAAGCCTAG 121
|||||
Db 171376 TGCACCTGAGAGACAGGACTGAGTTCCTTAGGCCGACTAAGATCCCTAAGCCTAG 171435

QY 122 CTGGGAAGGTGACCGTCCACTTTAAACGCGGCTTGCACTTAAGCTCAACCTGAC 181
|||||
Db 171436 CTGGGAAGGTGACCGTCCACTTTAAACGCGGCTTGCACTTAAGCTCAACCTGAC 171495

QY 182 CAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGACAGAGGTAAAGAAATAGCCAA 241
|||||
Db 171496 CAATCAGAGAGCTCACTAAATGCTAATTAGCAAAAGAGGTAAAGAAATAGCCAA 171555

QY 242 TCATCTAATTGCTGAGAGACAGCAGAGGAGACAAATGGGATTAACCCAGGCATT 301
|||||
Db 171556 TCATCTAATTGCTGAGAGACAGCAGAGGAGACAAATGGGATTAACCCAGGCATT 171615

QY 302 CGAGCTGGCAACAGCAGCCCCCTTTGGGGTCCCTTCCTTTGATAGGAG-CTGTTTTC 359
|||||
Db 171616 CGAGCTGGCAACAGCAGCCCCCTTTGGGGTCCCTTCCTTTGATAGGAGCTGTTTTC 171674

QY 360 ATGCTATTTCACCTCTAATTAATCTTGCAACTGACCTTCTGGTCCATGTTCTTAAGGC 419
|||||
Db 171675 ATGCTATTTCACCTCTAATTAATCTTGCAACTGACCTTCTGGTCCATGTTCTTAAGGC 171734

QY 420 TCGAGCTGAGCTTTTGCTACCGCTCAGCACTGCTGTTTGCCACCGCAGACCTGCGG 479
|||||
```

Db 171735 TCGAGCTGAGCTTTGCTCGCCATCCACACAGCTGTTTGCGGCTTGACAGACCGCTG 171794
Qy 480 CTGACTCCCATCCCTCTGGAT 500
Db 171795 CTGACTCCCATCCCTCTCGAT 171815

Search completed: April 19, 2003, 13:48:52
Job time : 2068.07 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:05:38 ; Search time 179.76 Seconds
(without alignments)
6263.892 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 ccctgggggggctccttctt.....tgactccatccctgcat 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	10499	21	ABN97929
2	500	100.0	56093	24	ABLe1744
3	435	87.0	711	21	ABN97946
4	413.4	82.7	7466	23	AA568626
5	410	82.0	783	20	AA25669
6	383.8	76.8	711	21	ABN97947
7	348.2	69.6	1329	19	AA43219
8	348.2	69.6	1329	20	AA29704
9	326.8	65.4	1393	22	AA531002

10	306.8	61.4	893	23	AA565964	DNA encoding novel
11	296.2	59.2	849	22	AA531000	Human diagnostic a
12	295.6	59.1	1478	23	AA587568	DNA encoding novel
13	292.4	58.5	808	23	AA588392	DNA encoding novel
14	290.8	58.2	1243	23	AA584189	DNA encoding novel
15	287.6	57.5	2629	22	AAK87491	Human immune/haema
16	286.4	57.3	2781	22	AA555630	Nucleotide sequenc
17	284.8	57.0	2946	20	AA577526	Human secreted pro
18	284.8	57.0	2946	21	AA594680	Human secreted pro
19	284.6	56.9	2306	23	AA592680	DNA encoding novel
20	283.8	56.8	2930	24	AA24195	Human synyrtin cDN
21	283.2	56.6	3372	20	AA525653	Human endogenous r
22	283.2	56.6	3372	21	AA559213	Partial pol gene a
23	283.2	56.6	7582	20	AA525655	Complete human end
24	283.2	56.6	7582	21	AA559215	Human endogenous r
25	282.2	56.4	3831	23	AA571727	DNA encoding novel
26	282.2	56.4	5154	23	AA567609	DNA encoding novel
27	282.2	56.4	8279	23	AA576474	DNA encoding novel
28	282.2	56.4	8294	23	AA584209	DNA encoding novel
29	281.6	56.3	2782	20	AA525661	Human endogenous r
30	281.6	56.3	2782	21	AA559211	5' non coding, 3'
31	281.6	56.3	2782	22	AA520069	HERV-W envelope pr
32	280.8	56.2	1155	23	AA572721	DNA encoding novel
33	280.6	56.1	1136	20	AA525660	Human endogenous r
34	280.6	56.1	1136	21	AA559210	3' pol gene and 3'
35	280	56.0	2942	23	AA577313	DNA encoding novel
36	279.2	55.8	17758	22	AAK84138	Human immune/haema
37	277.6	55.5	416	23	AA565963	DNA encoding novel
38	270.4	54.1	635	19	AA43215	Multiple sclerosis
39	270.4	54.1	635	20	AA529702	Clone CL6-3' from
40	270.4	54.1	2030	21	AA63826	Nucleotide sequenc
41	261.4	52.3	410	24	ABN94040	Gene #538 used to
42	261.4	52.3	410	24	ABK64210	Human benign prost
43	259.6	51.9	488	23	AA587567	DNA encoding novel
44	259.6	51.9	583	23	AA588391	DNA encoding novel
45	259.2	51.8	541	22	AA62368	Human foetal liver

ALIGNMENTS

RESULT 1
ABN97929
ID ABN97929 standard; DNA; 10499 BP.
XX
AC ABN97929;
XX
DT 01-AUG-2002 (first entry)
XX
DE Human retroviral sequence HERV-7g.
XX
KW Autoimmune disease; HERV-7g; chromosome 7g; immunotherapy;
KW multiple sclerosis; ds.
XX
OS Human retrovirus.
XX
PN WO9967395-A1.
XX
PD 29-DEC-1999.
XX
PF 23-JUN-1999; 99WO-FR01513.
XX
PR 23-JUN-1998; 98FR-0007920.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Alliel PM, Perin J, Rieger F;
XX
DR WPI; 2000-160587/14.
XX
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225pp; French.

XX The present invention relates to new nucleic acid sequences of human
XX endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
XX Regulatory elements associated with HERV-7q may alter expression of other
XX genes (even remote genes) on the same chromosome, inducing immunological
XX and/or neurological changes (which may be pathological or protective/
XX curative). HERV-7q peptides can be used to improve efficiency of the
XX immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
XX sequences can be used in immunogenic or vaccinating compositions, for
XX protection against autoimmune diseases, particularly multiple sclerosis.
XX The peptides may also be used (by sequence comparison) to detect/identify
XX endogenous retroviruses that are abnormally expressed in cancer,
XX neopathologies or other autoimmune diseases. The present sequence was
XX used to illustrate the invention.

XX Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;

Query Match 100.0%; Score 500; DB 21; Length 10499;

Best Local Similarity 100.0%; Pred. No. 5,1e-164; Mismatches 0; Indels 0; Gaps 0;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAACGCTGAGATACAGCAATTATC 60
DB 1 CCTTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAACGCTGAGATACAGCAATTATC 60
QY 61 TTGCACTGAGAGACAGAGCTAGCTGATTTCTTAGG TCGACTAAGAACTCTAAGCCTA 120
DB 61 TTGCACTGAGAGACAGAGCTAGCTGATTTCTTAGG TCGACTAAGAACTCTAAGCCTA 120
QY 121 GCTGGGAAGGTACACGCTCCACTTTTAAACAGGGGGTTGCAACTTAGCTACACCTGA 180
DB 121 GCTGGGAAGGTACACGCTCCACTTTTAAACAGGGGGTTGCAACTTAGCTACACCTGA 180
QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCA 240
DB 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCA 240
QY 241 ATCATCTATTGCTGAGACAGACAGAGGAGGAAACATCGGAGATTAACCCGAGCAT 300
DB 241 ATCATCTATTGCTGAGACAGACAGAGGAGGAAACATCGGAGATTAACCCGAGCAT 300
QY 301 TCGAGCTGGCAACAG 360
DB 301 TCGAGCTGGCAACAG 360
QY 361 TGGTATTGCTGATTAATCTTGGCACTGACTTTTGGTCCATGTTCTTAAGGCT 420
DB 361 TGGTATTGCTGATTAATCTTGGCACTGACTTTTGGTCCATGTTCTTAAGGCT 420
QY 421 CGAGCTGAGCTTTTGTCTACCGTCCACGACTGTTTTCACACCGAGAGAGAGAGAGAG 480
DB 421 CGAGCTGAGCTTTTGTCTACCGTCCACGACTGTTTTCACACCGAGAGAGAGAGAGAG 480
QY 481 TGAATCCCATCCCTCTGGAT 500
DB 481 TGAATCCCATCCCTCTGGAT 500

RESULT 2
ABL61744
ID ABL61744 standard; DNA; 56093 BP.

XX ABL61744;

XX 15-MAY-2002 (first entry)

XX Colon adenocarcinoma related gene sequence S3Q ID NO:81.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilms's tumour; adenocarcinoma;

KW gene; ds.

XX Homo sapiens.

XX WO200194629-A2.

XX 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US10838.

XX 05-JUN-2000; 2000US-209472P.

XX 05-JUN-2000; 2000US-209531P.

XX 18-SEP-2000; 2000US-231133P.

XX 18-SEP-2000; 2000US-231617P.

XX 20-SEP-2000; 2000US-234009P.

XX 20-SEP-2000; 2000US-234034P.

XX 20-SEP-2000; 2000US-234052P.

XX 22-SEP-2000; 2000US-234509P.

XX 22-SEP-2000; 2000US-234567P.

XX 25-SEP-2000; 2000US-234923P.

XX 25-SEP-2000; 2000US-234924P.

XX 25-SEP-2000; 2000US-235077P.

XX 25-SEP-2000; 2000US-235082P.

XX 25-SEP-2000; 2000US-235134P.

XX 25-SEP-2000; 2000US-235280P.

XX 26-SEP-2000; 2000US-235637P.

XX 26-SEP-2000; 2000US-235638P.

XX 27-SEP-2000; 2000US-235711P.

XX 27-SEP-2000; 2000US-235720P.

XX 27-SEP-2000; 2000US-235840P.

XX 27-SEP-2000; 2000US-235863P.

XX 28-SEP-2000; 2000US-236028P.

XX 28-SEP-2000; 2000US-236032P.

XX 28-SEP-2000; 2000US-236033P.

XX 28-SEP-2000; 2000US-236034P.

XX 28-SEP-2000; 2000US-236109P.

XX 28-SEP-2000; 2000US-236111P.

XX 29-SEP-2000; 2000US-236842P.

XX 29-SEP-2000; 2000US-236891P.

XX 02-OCT-2000; 2000US-237172P.

XX 02-OCT-2000; 2000US-237173P.

XX 02-OCT-2000; 2000US-237278P.

XX 02-OCT-2000; 2000US-237294P.

XX 02-OCT-2000; 2000US-237295P.

XX 03-OCT-2000; 2000US-237425P.

XX 03-OCT-2000; 2000US-237598P.

XX 03-OCT-2000; 2000US-237604P.

XX 03-OCT-2000; 2000US-237606P.

XX 03-OCT-2000; 2000US-237608P.

XX 01-NOV-2000; 2000US-244867P.

XX 01-NOV-2000; 2000US-245084P.

XX (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;

XX Sopet DR, Weaver Z;

XX WPI; 2002-188264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a

XX chemical agent to be tested for anti-neoplastic activity, and

XX determining a change in expression of a gene of a signature gene set

XX Claim 1; SEQ ID 81; 44pp; English.

XX The present invention describes a method (M1) for screening for an

XX anti-neoplastic agent. The method involves exposing cells to a chemical

XX agent to be tested for anti-neoplastic activity, determining a change in

XX expression of at least one gene (I) of a signature gene set, where (i)

XX comprises a sequence (S) selected from 8447 sequences (given in ABL6164

XX to ABL70110), or is at least 95% identical to (S), where a change in

XX expression is indicative of anti-neoplastic activity. (i) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC esophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;

Query Match 100.0%; Score 500; DB 24; Length 56093;
 Best Local Similarity 100.0%; Pred. No. 1.3e-163;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGTGGGGCGGGCTTCTCTTCTGGGATGAGGGGCAAAAGCCTGAGATACAGCAATTATC 60
 DB 28001 CCTGTGGGGCGGGCTTCTCTTCTGGGATGAGGGGCAAAAGCCTGAGATACAGCAATTATC 28060
 QY 61 TTGCACTGAGACAGAGATAGTGTGATTTCTTAGGCGCACTTAAGATCCCTTAAGCTTA 120
 DB 28061 TTGCACTGAGACAGAGATAGTGTGATTTCTTAGGCGCACTTAAGATCCCTTAAGCTTA 28120
 QY 121 GCTGGGAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGAACCTTCAACCTGA 180
 DB 28121 GCTGGGAGGTGACCAAGTCCACCTTTAAACAGGGGCTTGAACCTTCAACCTGA 28180
 QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGATTAAGAAATAGCCA 240
 DB 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGATTAAGAAATAGCCA 28240
 QY 241 ATATCTATTGCTCTGAGAGACAGAGAGAGAGACAAATCGGGATTAATACCAAGCAT 300
 DB 28241 ATATCTATTGCTCTGAGAGACAGAGAGAGAGACAAATCGGGATTAATACCAAGCAT 28300
 QY 301 TCGAGCTGAGACAG 360
 DB 28301 TCGAGCTGAGACAG 28360
 QY 361 TGTCTATTCACTCTAATTAATCTTGAACAGTCACTCTTGTGTCATGTTTTCAGGCT 420
 DB 28361 TGTCTATTCACTCTAATTAATCTTGAACAGTCACTCTTGTGTCATGTTTTCAGGCT 28420
 QY 421 CGAGCTGAGCTTTTGTCTCAACGTCACCACTGCTGTTTGGCAACGAGAGAGAGAGAGAG 480
 DB 28421 CGAGCTGAGCTTTTGTCTCAACGTCACCACTGCTGTTTGGCAACGAGAGAGAGAGAG 28480
 QY 481 TGAATCCCATCCCTCTGGAT 500
 DB 28481 TGAATCCCATCCCTCTGGAT 28500

RESULT 3

ABN97946
 ID ABN97946 standard; DNA; 711 BP.

XX AC ABN97946;
 XX 01-AUG-2002 (first entry)
 XX DE Human retroviral sequence R1.
 XX XX
 XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 XX multiple sclerosis; ds.
 XX Human retrovirus.
 XX XX
 XX PN WO967395-A1.
 XX PD 29-DEC-1999.
 XX XX

PF 23-JUN-1999; 99WO-FR01513.
 XX 23-JUN-1998; 98FR-0007920.
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX PI Alliel PM, Perin J, Rieger F;
 XX WPI; 2000-160587/14.

PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -
 PS Claim 3; Fig 3; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

XX Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;

Query Match 87.0%; Score 435; DB 21; Length 711;
 Best Local Similarity 100.0%; Pred. No. 7.1e-142;
 Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ACTGAGAGACAGAGACTAGTGTGATTTCTTAGGCGCACTAAGATCCCTAAGCTTACCTG 125
 DB 1 ACTGAGAGACAGAGACTAGTGTGATTTCTTAGGCGCACTAAGATCCCTAAGCTTACCTG 60
 QY 126 GAGGTGACCAAGTCCACTTTTAAACAGGGGCTTGCACCTTAGCTCACCTGACCAAT 185
 DB 61 GAGGTGACCAAGTCCACTTTTAAACAGGGGCTTGCACCTTAGCTCACCTGACCAAT 120
 QY 186 CAGAGAGCTCACTAAATGTAATTAAGGCAAGAGAGAGATTAAGAAATAGCAATCAT 245
 DB 121 CAGAGAGCTCACTAAATGTAATTAAGGCAAGAGAGATTAAGAAATAGCAATCAT 180
 QY 246 CTATTGCTGAGAGACAG 305
 DB 181 CTATTGCTGAGAGACAG 240
 QY 306 CTGGCAACAG 365
 DB 241 CTGGCAACAG 300
 QY 366 TTTCACCTAATTAATCTTGAACAGTCACTCTGAGTCAATGTTTCTTAAGGCTGAGC 425
 DB 301 TTTCACCTAATTAATCTTGAACAGTCACTCTGAGTCAATGTTTCTTAAGGCTGAGC 360
 QY 426 TGAAGCTTTTGTCTCAACGTCACCACTGCTGTTTGGCAACCGAGAGAGAGAGAGAG 485
 DB 361 TGAAGCTTTTGTCTCAACGTCACCACTGCTGTTTGGCAACCGAGAGAGAGAGAGAG 420
 QY 486 CCCATCCCTCTGGAT 500
 DB 421 CCCATCCCTCTGGAT 435

RESULT 4

AAS68626
 ID AAS68626 standard; cDNA; 7466 BP.

AC AAS68626;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #4430.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX MO200175067-A2.
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX P-PSDB; ABG04439.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX
XX Claim 1; SEQ ID No 4430; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1'54 T; 1 other;
XX
XX
XX Query Match 82.7%; Score 413.4; LB 23; Length 7466;
XX Best Local Similarity 98.2%; Pred. No. 1e-133;
XX Matches 429; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

Oy 246 CTATTGCTGAGACAGCAGAGGACAAACATCGGATATAAACCAGCATTCGAG 305
Db 3896 TTATTGCTGAGACAGCAGAGGACAAATGATCGGATATAAACCAGCATTCGAG 3955
Oy 306 CTGCAACAGCAGGCCCCCTTTGGGTCCTTCCCTTTGATGGAG--CTGTTTCATGC 363
Db 3956 CTGCAACAGCAGGCCCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCATGC 4015
Oy 364 TATTCACTCATTAATCTTGAACATGCACTCTTGGTCAGATTTCTTAACGGCTCGA 423
Db 4016 TATTCACTCATTAATCTTGAACATGCACTCTTGGTCAGATTTCTTAACGGCTCGA 4075
Oy 424 GCTGAGCTTTTGTACCGTCACACCACTGCTGTGTGACACCGCAGACCTGCGCTGA 483
Db 4076 GCTGAGCTTTTGTACCGTCACACCACTGCTGTGTGACACCGCAGACCTGCGCTGA 4135
Oy 484 CTCCCATCCTCTTGGAAT 500
Db 4136 CTCCCATCCTCTTGGAAT 4152

RESULT 5
AAX25669
ID AAX25669 standard; cDNA to mRNA; 783 BP.

AAX25669;

XX 21-MAY-1999 (first entry)

XX Human endogenous retrovirus W long terminal repeat region.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;
XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX disseminated lupus erythematosus; pregnancy; chromosome marker; ss.

XX Human endogenous retrovirus.

XX MO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (IMR) BIO MERIEUX.

XX Beeseme F, Blond JL, Boulton O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
XX expressed exclusively in placenta and useful in diagnosis and
XX therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 83; 106bp; French.

XX This sequence represents the long terminal repeat (LTR) region of the
XX human endogenous retrovirus (HERV) W genome. The nucleic acids, their
XX fragments or peptides encoded by them are markers of autoimmune disease
XX (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus
XX erythematosus, insulin-dependent diabetes and related pathologies) and
XX of abnormal or unsuccessful pregnancy and can be used as chromosomal
XX markers for susceptibility to these conditions, or proximity markers
XX of genes associated with this susceptibility.

XX Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;

XX Query Match 82.0%; Score 410; DB 20; Length 783;
XX Best Local Similarity 93.8%; Pred. No. 4.5e-133;
XX Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;

QY 68 TGAAGACAGAGCTAGCTGATTTCTTAGGCGCACTAAGAAATCCCTTAGCTAGCTGAGG 127
 CC |||||
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

QY 128 AGGTGACGACGCTGCTTTAAACGCGGCTTGCACCTTAGCTCACAACCTGACCAATCA 187
 CC |||||
 CC 61 AGGTGACGACGCTGCTTTAAACGCGGCTTGCACCTTAGCTCACAACCTGACCAATCA 120
 CC |||||
 CC 188 GAGAGCTCAGTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCAATCATCT 247
 CC |||||
 CC 121 GAGAGCTCAGTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCAATCATCT 180
 CC |||||
 CC QY 248 ATTGCTTGAAGACAGACGAGGAGGACACATCGGATATATAACCAAGCATTCAGACT 307
 CC |||||
 CC Db 181 ATTGCTTGAAGACAGACGAGGAGGACACATCGGATATATAACCAAGCATTCAGACT 240
 CC |||||
 CC QY 308 GGCACAGCAGGCCCCCTTTGGGTCCTTCCCTTGTATGGAG--CTGTTTCAATGCTA 365
 CC |||||
 CC Db 241 GGCACAGCAGGCCCCCTTTGGGTCCTTCCCTTGTATGGAGCTCTGTTTCAATGCTA 300
 CC |||||
 CC QY 366 TTTCACCTTAATTAATCTTGCACTGCACTGCTGTCATGTTTCTTAAGGCTGAGC 425
 CC |||||
 CC Db 301 TTTCACCTTAATTAATCTTGCACTGCACTGCTGTCATGTTTCTTAAGGCTGAGC 360
 CC |||||
 CC QY 426 TGAAGCTTTGCTCAGCGTCCACACTGCTGTTTGCACCAACCGGAGACTGCGCTGACT 485
 CC |||||
 CC Db 361 TGAAGCTTTGCTCAGCGTCCACACTGCTGTTTGCACCAACCGGAGACTGCGCTGACT 420
 CC |||||
 CC QY 486 CCGATCCCTCTGGAT 500
 CC |||||
 CC Db 421 CCGATCCCTCTGGAT 435
 CC |||||

RESULT 6

ABN97947
 ID ABN97947 standard; DNA; 711 BP.
 XX
 AC ABN97947;
 XX
 XX

DT 01-AUG-2002 (first entry)
 XX
 DE Human retroviral sequence R1F.
 XX
 XX

Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 multiple sclerosis; ds.
 XX
 XX

Human retrovirus.
 OS
 XX

WO967395-A1.
 PN
 XX

29-DEC-1999.
 PD
 XX

23-JUN-1999; 99MO-FR01513.
 PF
 XX

23-JUN-1998; 98FR-0007920.
 PR
 XX

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA
 XX

Alliel PM, Perin J, Rieger F;
 PI
 XX

WPI; 2000-160587/14.
 DR
 XX

New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -
 PT

neurological diseases -
 PS
 XX

Claim 3; Fig 3; 225pp; French.
 PS
 XX

The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/

CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

SO Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;

Query Match 76.8%; Score 383.8; DB 21; Length 711;
 Best Local Similarity 95.4%; Pred. No. 6.6e-124;
 Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAGACAGAGCTAGCTGATTTCTTAGGCGCACTAAGAAATCCCTTAGCTAGCTGAG 125
 CC |||||
 CC Db 1 ACTGAGACAGAGCTAGCTGATTTCTTAGGCGCACTAAGAAATCCCTTAGCTAGCTGAG 60
 CC |||||
 CC QY 126 GAAGTGCACGCTCCACCTTTAAACGCGGCTTGCACCTTAGCTCACAACCTGACCAAT 185
 CC |||||
 CC Db 61 GAAGTGCACGCTCCACCTTTAAACGCGGCTTGCACCTTAGCTCACAACCTGACCAAT 120
 CC |||||
 CC QY 186 CAGAGAGCTCAGTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCAATCAT 245
 CC |||||
 CC Db 121 CAGAGAGCTCAGTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCAATCAT 180
 CC |||||
 CC QY 246 CTATGCTTGAAGACAGACGAGGAGGACACATGCGGATATAACCCAGCATTCGAG 305
 CC |||||
 CC Db 181 CTATGCTTGAAGACAGACGAGGAGGACACATGCGGATATAACCCAGCATTCGAG 240
 CC |||||
 CC QY 306 CTGCAACAGCAGCCCCCTTTGGGTCCTTCCCTTGTATGGAG--CTGTTTCAATGCT 363
 CC |||||
 CC Db 241 CTGCAACAGCAGCCCCCTTTGGGTCCTTCCCTTGTATGGAGCTCTGTTTCAATGCT 299
 CC |||||
 CC QY 364 TATTTCACTATTAATCTTTGCACTGCACTGCTGTCATGTTTCTTAAGGCTGCA 423
 CC |||||
 CC Db 300 TATTTCACTATTAATCTTTGCACTGCACTGCTGTCATGTTTCTTAAGGCTGCA 359
 CC |||||
 CC QY 424 GCTGAGCTTTGCTCAGCGTCCACACTGCTGTTTGCACCAACCGGAGACTGCGCTGCA 483
 CC |||||
 CC Db 360 GCTGAGCTTTGCTCAGCGTCCACACTGCTGTTTGCACCAACCGGAGACTGCGCTGCA 419
 CC |||||
 CC QY 484 CTCCCATCCCTCTGGAT 500
 CC |||||
 CC Db 420 CTCCCATCCCTCTGGAT 436
 CC |||||

RESULT 7

AAV43219
 ID AAV43219 standard; cDNA; 1329 BP.
 XX
 AC AAV43219;
 XX
 XX

29-DEC-1998 (first entry)
 DT
 XX

Multiple sclerosis associated retrovirus fragment 6.
 DE
 XX

Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
 KW gag gene; env gene; rheumatoid arthritis-associated virus; ss.
 KW
 XX

Multiple sclerosis associated retrovirus.
 OS
 XX

Location/Qualifiers
 FH
 FT

Key
 FT
 CDS

2..490
 /tag= a
 /product= "Encodes protein AAV71069"
 /transl_except= (pos:77-79, appears to code for a
 stop codon)
 /transl_except= (pos:125-127, appears to code for a
 stop codon)
 /transl_except= (pos:137-139, appears to code for a
 stop codon)

```
XX MO9823755-A1.
PN
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-IB01482.
XX
XX 26-NOV-1996; 96US-0756429.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Bedin F, Beeseme F, Jolivet-Reynaud C, Komurian-Pradel F,
PI Mandrand B, Paranhos-Baccala G, Perron J;
XX
XX WPI: 1998-332732/28.
XX
XX P-PSDB; AAW71069.
XX
XX New nucleic acid from retroviruses - useful for diagnosis,
PT prevention and treatment of, e.g. multiple sclerosis
XX
XX Disclosure; Pages 187-188; 286pp; English.
XX
XX The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) genomic fragment used in the method of the
CC invention. The invention provides complete or partial genomic
CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
CC polypeptides encoded by these genes. The invention also provides
CC antibodies raised against the polypeptides. The genomic sequences,
CC polypeptides and antibodies are also claimed useful for diagnosing
CC infection by MS and rheumatoid arthritis-associated viruses, and also
CC for prevention and treatment of infection with these viruses.
XX
SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;
Query Match 69.6%; Score 348.2; DB 19; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;
QY 66 ACTGAGAGACAGGACTAGTGATTTCTTAGGCGACCAAGAAATCCTAAGCCTAGCTGG 125
DB 547 ACTGAGAGACAGGACTAGTGATTTCTTAGGCTGACTAAGAAATCCNAAGCCTANCTGG 606
QY 126 GAAGGTGACACGTCACCTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCTTAAATGCAATGGGGCTTGCACTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGAGGTAAAGCAATAGCCATCAT 726
QY 246 CTATTGCTGAGAGACAGCAGAGGAGCAACAATCGGATATAAACCAGGACTTGGAG 305
DB 727 CTATTGCTGAGAGACAGCAGGAGGAGCAAGAAATTTGGATATAAATCAGGACTTCAAG 786
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCTTTGTATGGAGAG--CTGTTTTATGTC 363
DB 787 CCAGCAACAGCA-ACCCCTTTGGGTCCCTTCCTTCATTTGATGGAGAGCTGTTTACATC 845
QY 364 TATTTCACCTAATTAATTTTGAACATGCACTCTTCTGATTCATGTTTCTTAAGGCTGA 423
DB 846 TATTTCACCTAATTAATGATGCAACTGCACTCTTCTGATTCGATGTTTATAGGCTCAA 905
QY 424 GGTGAGCTTTTGTCTACACGCTGACCACTGCTGTTGGCCCAACCGGAGACTGCGGCTGA 483
DB 906 GGTGAGCTTTTGTGCGCATCCACCACTGCTGTTGGCCCAACCGGAGACTGCGGCTGA 965
QY 484 CTCCCATCCCTCTGGAT 500
DB 966 CTTCATCCCTTTGGAT 982
RESULT 8
AAK29704
```

```
ID AAK29704 standard; DNN; 1329 BP.
XX
XX AAK29704;
AC
XX
XX 08-JUN-1999 (first entry)
DT
XX
XX Clone 5M6 from MSRV-1.
DE
XX
XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KM rheumatoid polyarthritis; ss.
XX
XX OS Multiple sclerosis related virus type 1.
XX
XX PN FR2765588-A1.
XX
XX PD 08-JAN-1999.
XX
XX PF 07-JUL-1997; 97FR-0008816.
XX
XX PR 07-JUL-1997; 97FR-0008816.
XX
XX (INMR ) BIO MERIEUX.
XX
XX WPI: 1999-098275/09.
XX
XX P-PSDB; AAW95554.
XX
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis
XX
XX PS Claim 1; Page 39-40; 83pp; French.
XX
XX This sequence represents clone 5M6 from a novel multiple sclerosis
CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with
CC rheumatoid polyarthritis.
XX
SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;
Query Match 69.6%; Score 348.2; DB 20; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;
QY 66 ACTGAGAGACAGGACTAGTGATTTCTTAGGCGGACTAAGAAATCCTAAGCCTAGCTGG 125
DB 547 ACTGAGAGACAGGACTAGTGATTTCTTAGGCTGACTAAGAAATCCNAAGCCTANCTGG 606
QY 126 GAAGGTGACACGTCACCTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCTTAAATGCAATGGGGCTTGCACTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTAATCAGGCAAAAACAGAGGTAAAGCAATAGCCATCAT 726
QY 246 CTATTGCTGAGAGACAGCAGAGGAGCAACAATCGGAGATATAAACCAGGACTTGGAG 305
DB 727 CTATTGCTGAGAGACAGCAGGAGGAGCAAGAAATTTGGATATAAATCAGGACTTCAAG 786
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCTTTGTATGGAGAG--CTGTTTTATGTC 363
DB 787 CCAGCAACAGCA-ACCCCTTTGGGTCCCTTCCTTCATTTGATGGAGAGCTGTTTACATC 845
QY 364 TATTTCACCTAATTAATTTTGAACATGCACTCTTCTGATTCATGTTTCTTAAGGCTGA 423
DB 846 TATTTCACCTAATTAATGATGCAACTGCACTCTTCTGATTCGATGTTTATAGGCTCAA 905
QY 424 GGTGAGCTTTTGTCTACACGCTGACCACTGCTGTTGGCCCAACCGGAGACTGCGGCTGA 483
DB 906 GGTGAGCTTTTGTGCGCATCCACCACTGCTGTTGGCCCAACCGGAGACTGCGGCTGA 965
QY 484 CTCCCATCCCTCTGGAT 500
```

Db 966 CTTCCATCCTTGAT 982

RESULT 9

AAS31002 standard; cDNA; 1393 BP.

AAS31002;

04-DEC-2001 (first entry)

Human diagnostic and therapeutic polynucleotide (DITHP) #17.

Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
acquired immune deficiency syndrome; AIDS; autoimmune disorder;
respiratory disorder; ss.

Homo sapiens.

WO200162927-A2.

30-AUG-2001.

21-FEB-2001; 2001WO-US06059.

24-FEB-2000; 2000US-0184693.

24-FEB-2000; 2000US-0184697.

24-FEB-2000; 2000US-0184598.

24-FEB-2000; 2000US-0184768.

24-FEB-2000; 2000US-0184770.

24-FEB-2000; 2000US-0184772.

24-FEB-2000; 2000US-0184773.

24-FEB-2000; 2000US-0184776.

24-FEB-2000; 2000US-0184777.

24-FEB-2000; 2000US-0184797.

24-FEB-2000; 2000US-0184813.

24-FEB-2000; 2000US-0184837.

24-FEB-2000; 2000US-0184841.

24-FEB-2000; 2000US-0185213.

12-MAY-2000; 2000US-0203785.

15-MAY-2000; 2000US-0204226.

16-MAY-2000; 2000US-0204525.

16-MAY-2000; 2000US-0204821.

16-MAY-2000; 2000US-0204908.

17-MAY-2000; 2000US-0205332.

17-MAY-2000; 2000US-0204815.

17-MAY-2000; 2000US-0205221.

17-MAY-2000; 2000US-0205285.

17-MAY-2000; 2000US-0205286.

17-MAY-2000; 2000US-0205323.

17-MAY-2000; 2000US-0205324.

(INCY-) INCYTE GENOMICS INC.

Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;

Chen A, D'Sa SA, Ameshey S, Dahl CR, Dam TC, Daniels SE;

Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;

Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreiter TK, Daffo A;

Wright RJ, Yap PB, Yu JY, Bradley DL, Bratcher SR, Chen W;

Cohen HD, Hodgson DM, Lincoln SE, Jackson SJ;

WPI; 2001-502867/55.

P-PSDB; AAU19431.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g. enzymes, hormones and receptors, useful in diagnostics and therapeutics

PT -
XX Claim 1; Page 304; 522pp; English.
XX

CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHP, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHP, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbent
CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
CC therapeutic (DITHP) polynucleotides of the invention.

SQ Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 2 other;

Query Match 65.4%; Score 326.8; DB 22; Length 1393;

Best Local Similarity 88.7%; Pred. No. 16-103;

Matches 394; Conservative 0; Mismatches 32; Indels 18; Gaps 3;

66 ACTGAGAGACAGGACTGATGATTTCTAGGCGGACTAAGATCCCTAGCTAGCTG 125

245 ACTGAGAGACAGGACTGATGATTTCTAGGCGGACTAAGATCCCTAGCTAGCTG 304

126 GAGGTGACCACTGCTTAAACAGGGGGCTTGCAATTAGCTACCTGACCAAT 185

305 GAGGTGACCGCTCTTAAACAGGGGGCTTGCAATTAGCTACCAAT 364

186 C-----AGAGACTGACTAAATGCTAAATAGGCAAGACAGAGATAAGAAAT 236

365 CAGGTAGTAAGAGACTGCTAAATGCTAAATAGGCAAGACAGAGATAAGAAAT 424

237 GCCAATCATCTATTGCTGAGACAGACAGAGAGGACAAATCGGATATTAACCCAG 296

425 GCCAATCATCTATTGCTGAGACAGACAGAGAGGACAAATCGGATATTAACCCAG 484

297 GCATTGAGCTGGCAACAGAGCCCTTTGGGTCCTTCTTGTATGAGAGCTGTT 356

485 GCATTGAGCTGGCAACAGAGCCCTTTGGGTCCTTCTTGTATGAGAGCTGTT 543

357 TTGATGCTATTGCTGCTTAAATCTGCACTGACCTTCTGTCATGTTCTTAC 416

544 T-----TTTCACTCTATTAATCTTGAACCTGCACTTCTGTTGTTGTTAC 595

417 GAGCTGAGCTGAGCTTTTCTACCGCTCACACTGCTTTTGCCACCAACCGAGACTG 476

596 GAGTTGAGCTGAGCTTTTCTACCGCTCACACTGCTTTTGCCACCAACCGAGACTG 655

477 CCGCTGACTCCCATCCCTCTGAT 500

656 CCGCTGACTCCCATCCCTCTGAT 679

WPI; 2001-502867/55.

P-PSDB; AAU19431.

Polynucleotides encoding diagnostic and therapeutic proteins, e.g. enzymes, hormones and receptors, useful in diagnostics and therapeutics

AC AAS65964;
XX
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #1768.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Dmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG01777.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 1768; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products represent novel human
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electron.c format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;
Query Match 61.4%; Score 306.8; IB 23; Length 893;
Best Local Similarity 87.2%; Pred. No. 8.1e-97;
Matches 389; Conservative 0; Mismatches 37; Indels 20; Gaps 4;
QY 66 ACTGAGAGCAGAGCTAGTGAATTTCTTAGGCGGACTAGATCCCTAGGCTAGCTGG 125
DB 236 ACTGAGAGCAGAGCTAGTGAATTTCTTAGGCGGACTAGATCCCTAGGCTAGCTGG 295
QY 126 GAAGGTGACGACGCTTAAACACCGGGCTTGCACTTAGCTCACACCTGACCAT 185
DB 296 GAAGGTGACGACGCTTAAACACCGGGCTTGCACTTAGCTCACACCTGACCAT 355
QY 186 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAGAGCAGAGGTAAAGAAATA 236
DB 356 CAGGTAAAGAAAGAGCCCGCTAAATGCTAATTAGGCAAGAGCAGAGGTAAAGAAATA 415

QY 237 GCCATCATCTATTTGCTTGAGACACAGACAGAGGACAAATCGGGATATAACCCAG 296
DB 416 GCCATCATCTATTTGCTTGAGACACAGACAGAGGAGGACAAATGATATTAACCCAG 475
QY 297 GCATTGAGCTGACACAGAGAGCCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTT 356
DB 476 GCATTGAGCTGACACAGAGAGCCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTG 534
QY 357 TTCATGCTATTTCACTCTATTAATCTTGCACTG--CACTCTTCTGTCATGTTCTT 414
DB 535 T-----TTTCACTCTATTAATCTTGCACTGCACTCTTCTGTCATGTTCTT 586
QY 415 ACCGCTGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTTGGCACCAGACCC 474
DB 587 ACCGCTGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTTGGCACCAGACCC 646
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
DB 647 CATGCTGACTTCCACCCCTCCAGAT 672
RESULT 11
AAS31000
ID AAS31000 standard; cDNA; 849 BP.
XX
AC AAS31000;
XX
DT 04-DEC-2001 (first entry)
XX
XX Human diagnostic and therapeutic polynucleotide (DITHP) #15.
DE
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KM cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
KM acquired immune deficiency syndrome; AIDS; autoimmune disorder;
XX respiratory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200162927-A2.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001MO-US06059.
XX
XX 24-FEB-2000; 2000US-0184693.
PR 24-FEB-2000; 2000US-0184697.
PR 24-FEB-2000; 2000US-0184698.
PR 24-FEB-2000; 2000US-0184768.
PR 24-FEB-2000; 2000US-0184769.
PR 24-FEB-2000; 2000US-0184770.
PR 24-FEB-2000; 2000US-0184771.
PR 24-FEB-2000; 2000US-0184772.
PR 24-FEB-2000; 2000US-0184773.
PR 24-FEB-2000; 2000US-0184774.
PR 24-FEB-2000; 2000US-0184776.
PR 24-FEB-2000; 2000US-0184777.
PR 24-FEB-2000; 2000US-0184797.
PR 24-FEB-2000; 2000US-0184813.
PR 24-FEB-2000; 2000US-0184837.
PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204525.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 17-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.

PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX
XX
XX (INCYTE) INCYTE GENOMICS INC.
XX
XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'Sa SN, Amesby S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenawald LB, Hillman JL, Jones AL;
PI Liu YT, Roseberry AM, Rosen BH, Russo PD, Stockdreher TK, Daffio A;
PI Wright RJ, Yap PE, Yu JY, Bredley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
XX WPI: 2001-502867/55.
XX P-PSDB; AAU19429.
XX
XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
XX PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
XX
XX
XX Claim 1; Page 303; 522pp; English.

The invention relates to polynucleotides (I) encoding diagnostic and therapeutic (DlTRHP) polypeptides (II), which include e.g. enzymes, CC and proteins involved in growth and development and receptors. (I) and CC (II) may be used in the prevention, diagnosis and treatment of diseases CC associated with inappropriate DTRHP expression. For example, (I) and CC (II) may be used to treat disorders associated with decreased polypeptide CC expression by rectifying mutations or deletions in a patient's genome, CC that affect the activity of the DTRHPs, by expressing inactive proteins CC or supplementing the patient's own production of them. (I) and (II) CC may be used to treat diseases, for example, cell proliferative disorder, CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma, CC leukemia, autoimmune disorders, and respiratory disorders. Additionally, CC (I) may be used to produce the DTRHPs, by inserting the nucleic acids CC into a host cell and culturing the cell to express the protein. (I) and CC its complementary sequences may also be used as DNA probes in diagnostic CC assays to detect and quantitate the presence of similar nucleic acids in CC samples, and therefore which patients may be in need of restorative CC therapy. (II) may also be used as antigens in the production of CC antibodies against DTRHPs and in assays to identify modulators of DTRHP CC expression and activity. The anti-DTRHP antibodies and antagonists may CC also be used to down regulate expression and activity. The anti-DTRHP CC antibodies may also be used as diagnostic agents for detecting the CC presence of DTRHPs in samples (e.g. by enzyme linked immunosorbent CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and CC therapeutic (DlTRHP) polynucleotides of the invention. CC

Sequence 849 BP, 233 A, 232 C, 184 G, 200 T, 0 other, CC

Query Match	59.2%	Score 296.2	DB 22	Length 849
Best Local Similarity	85.5%	Pred No. 4.1e-93		
Matches 382	Conservative	0	Mismatches 48	Indels 17
				Gaps 4

OY	65	AACGAGAGACAGGACTAGCGGATTTCTTGAGCCGATAAGAAATCCTTAAGCCTAGCTG	123
Db	384	AACTAGAGACAGGACTAGCGGATTTCTTGAGCCGATAAGAAATCCTTAAGCCTAGCTG	443
OY	125	GGAAGGTACCACTGTCACCTTTAAACAGGGGGCTTGGAACTTAGCTACACCTGACCA	184
Db	444	GGAGGTACCGGTATCCACCTTTAAACATAGGGCTTGGAACTTAGCTACACCTGACCTAA	503
OY	185	TCA-----GAGAGCTCACTAAATGCTTAATTAGGCAAGAAGAGGCTAAAGAAAT	235
Db	504	TCAGATAGTAAGAGAGGCTCACTAAATGCTTAATTAGGCAAAAAAGAGGCTAAAGAAAT	563
OY	236	AGCCATCATCTTAATGGCCTGAGAGACAGCAGAGAGGACCAATCGGATATATAACCA	295
Db	564	AGCCATCATCTTAATGGCCTGAGAGACAGCAGAGAGGACCAATGATCAGATATATAACCA	622
OY	236	GGCATTGAGCTGGCAACAGACAGCCCCCTTTGGATCCTTCCCTTTGTATGGGAG--CT	353

Db	624	GGCATTCAAGCAGACANTGGC-TACCTCTTTGGGTCCCTCCGTTTGATGGAGACTCT	682
Oy	354	GTTTTCATGCTATTCTACTCTATTAAATCTTGCACTGCACTCTTCTGGTCATGTTTCT	413
Db	683	GTTTTCATGCTATTTTAAAGCTTGCAACT-----GCACACTCTTCTGATCCATGTTGT	727
Oy	414	TACGGCTGAGACTAGGTTTTGTGCACACGTCACCATGTGCTGTTTTGGCAACACCGAGAC	473
Db	738	TACAGCTTGAGCTAGGCTTTTGTGCACACATCCAGCAAMGTCTTTGGCAACCATCAGAC	797
Oy	474	CTGCGCTGACTCCCATCCCTCTGGAT	500
Db	798	CCGCGACTGACTTTCATCCCTCTGGAT	824

RESULT 12
AAS87568/c
ID AAS87568 standard; cDNA; 1478 BP

AC AAS87568;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #23372.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1; SEQ ID No 23372; 103bp; English

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. Ab564197-Ab594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1478 BP; 389 A; 337 C; 364 G; 383 T; 0 other;
SQ

Query Match 59.1%; Score 295.6; DB 23; Length 1478;
Best Local Similarity 85.7%; Pred. No. 9.2e-93;
Matches 382; Conservative 0; Mismatches 44; Indels 20; Gaps 4;

QY 66 ACTGAGACAGGAGTCTGATTTCTAGGCGCAATTAAGAAATCCCTAAGCTTACCTG 125
DB 1473 ATTGAAACAGGAGCAGCTGATCTCTAGGCTGACTTAATAATCTTAAGCTTACCTG 1414
QY 126 GAAGGTGACCAAGTCCACTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAAT 185
DB 1413 GAAGGTGACTGATCCACTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAAT 1354
QY 186 C-----AGAGGCTCACTAAATGCTAATTGCGAAAGACAGGAGTTAAGAAATA 236
DB 1353 CAGGTAGTAAAGAGAGCTCACTAAACGCTCATTTAGCAAAAACAGGAGTTAAGAAATA 1294
QY 237 GCCATCATCTATTGCGTGAAGACAGAGAGGAGCAACAATGGGATTAATTAACCCAG 296
DB 1293 GCCATCATCTATTGCTGAGAGCAAGAGGAGGAGGAATGACCCGGAATTTAAACCCAG 1234
QY 297 GCATTGAGCTGCAACAGACAGCCCTTGGGCTCCCTTCCCTTTGATGAGAGCTGTT 356
DB 1233 GCATTGAGCTGCAACAGGCTTACCTCTCTGGGCTCACTCCCTTTGATGAGAGCTG 1175
QY 357 TTGATGCTATTGCTGCTATTAAATCTTGCAACTG--CACTCTTGTGCTCATGTTTCTT 414
DB 1174 T-----TTTCACTCTATTAAATCTTGCAACTGCAAACTCTTGTGCTCATGTTTCTT 1123
QY 415 ACGGCTCAGCTGAGCTTTTGTCTACCGTCCAGCACTCTGTTTGCAACACCGAGACC 474
DB 1122 ATGGCTCAGCTGAGCTTTGCTGACAGTCCACACTCTCTGTTTCCCGCATGCAAGCC 1063
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
DB 1062 CGTACTGACTTCCACCCCTCCGGAT 1037

RESULT 13
AAS88392
ID AAS88392 standard; cDNA; 808 BP.
XX
AC AAS88392;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #24196.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
AC P-PSDB; ABG24205.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 24196; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 808 BP; 184 A; 234 C; 189 G; 201 T; 0 other;
Query Match 58.5%; Score 292.4; DB 23; Length 808;
Best Local Similarity 85.2%; Pred. No. 8.6e-92;
Matches 380; Conservative 0; Mismatches 46; Indels 20; Gaps 4;

QY 66 ACTGAGACAGGAGTCTGATTTCTAGGCGCAATTAAGAAATCCCTAAGCTTACCTG 125
DB 6 ATTGAAACAGGAGCAGCTGATCTCTAGGCTGACTTAAGAAATCCCTAAGCTTACCTG 65
QY 126 GAAGGTGACCAAGTCCACTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAAT 185
DB 66 GAAGGTGACTGATCCACTTTAAACAGGGGCTTGCACTTAGCTCAACCTGACCAAT 125
QY 186 C-----AGAGGCTCACTAAATGCTAATTAGGCAAAAGACAGAGTTAAGAAATA 236
DB 126 CAGGTAGTAAAGAGAGCTCACTAAACGCTCATTTAGCAAAAACAGGAGTTAAGAAATA 185
QY 237 GCCATCATCTATTGCTGAGAGACAGACAGGAGGAGCAACAATGGGATTAATTAACCCAG 296
DB 186 GCCATCATCTATTGCTGAGAGACAGGAGGAGGAGCAACAATGGGATTAATTAACCCAG 245
QY 297 GCATTGAGCTGCAACAGACAGCCCTTGGGCTCCCTTCTTGTGATGAGAGCTGTT 356
DB 246 GCATTGAGCTGCAACAGGCTTACCTCTCTGGGCTCACTCTTGTGATGAGAGCTG 304
QY 357 TTGATGCTATTGCTGCTATTAAATCTTGCAACTG--CACTCTTGTGCTCATGTTTCTT 414
DB 305 T-----TTTCACTCTATTAAATCTTGCAACTGCAACTCTTGTGCTCATGTTTCTT 356
QY 415 ACGGCTCAGCTGAGCTTTTGTCTACCGTCCAGCACTGCTTTTGCAACACCGAGACC 474
DB 357 ACAGCTTAGCTGAGCTTTCACTGCGGCTCCACACTGCTGTTGTGCGCGTCAAGACC 416
QY 475 TGCCGCTGACTCCCATCCCTCTGGAT 500
DB 417 CACGCTGACTTCCACCCCTCCGGAT 442

RESULT 14
AAS84189/c
ID AAS84189 standard; cDNA; 1243 BP.
XX
AC AAS84189;
XX

DT	13-FEB-2002	(first entry)
XX		
DE	DNA encoding novel human diagnostic protein #19993.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Dzmannac RT, Liu C, Tang YT,	
XX		
DR	WPI; 2001-639362/73.	
DR	P-PSDB; ABG20002.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity -	
XX		
PS	Claim 1; SEQ ID No 19993; 103pp; English.	
XX		
CC	The invention relates to isolated polynucleotide (I) and	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
CC	and gene mapping, and in recombinant production of (II). The	
CC	polynucleotides are also used in diagnostics as expressed sequence tags	
CC	for identifying expressed genes. (I) is useful in gene therapy techniques	
CC	to restore normal activity of (II) or to treat disease states involving	
CC	(II). (II) is useful for generating antibodies against it, detecting or	
CC	quantitating a polypeptide in tissue, as molecular weight markers and as	
CC	a food supplement. (II) and its binding partners are useful in medical	
CC	imaging of sites expressing (II). (I) and (II) are useful for treating	
CC	disorders involving aberrant protein expression or biological activity.	
CC	The polypeptide and polynucleotide sequences have applications in	
CC	diagnostics, forensics, gene mapping, identification of mutations	
CC	responsible for genetic disorders or other traits to assess biodiversity	
CC	and to produce other types of data and products dependent on DNA and	
CC	amino acid sequences. AAS64197-AAS94564 represent novel human	
CC	diagnostic coding sequences of the invention.	
CC	Note: The sequence data for this patent did not appear in the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp://wipo.int/pub/published_pct_sequences .	
XX		
SO	Sequence 1243 BP; 285 A; 334 C; 358 G; 266 T; 0 other;	
	Query Match	58.2%; Score 290.8; DB 23; Length 1243;
	Best Local Similarity	85.0%; Pred. No. 4e-91;
	Matches	379; Conservative 0; Mismatches 47; Indels 20; Gaps 4;
QY	66 ACTAGAGACAGGACTTAGTGTGATTTCTTAGGCGGACTTAAGATCCCTTAAGCCTTAGCTGG	125
DB	1238 ATTGAAACACAGGACAGCTGTGATCTCTTAGGCTTAAGATCCCTTAAGCCTTAGCTGG	1179
QY	126 GAAGGTGACACAGCTCCACTTTAAACAGGGGCTTGCAACTTAAGCTCACACTGACCAAT	185
DB	1178 GAAAGTGAAGTGCATCACTTTAAACAGGGGCTTGCAACTTAAGCTCACACCCGACCAAT	1119
QY	186 C-----AGAGAGCTCACTTAAATGCTTAATTTAGGCAGCAAGAGAGTAAAGAAATA	236
DB	1118 CAGGTAAAGTAAAGAGAGCTCACTTAAACGCTCACTTTAGGCAGCAAGAGAGTAAAGAAATA	1059
QY	237 GCCAATCATCTATTGCTGAGAGCACAGCAGAGGAGCAACAATGGGATTAATTAACCCAG	296

Db	1058	GCCATCATCTATTGTCTGAGAGACACCGGAGGAGCAATGACCCGGATTATAACCCAG	9939
Qy	297	GCATTGAGCTGGACAGACAGACCCCCCTTGGGTCCCTTCCCTTTGATGGAGCTGT	356
Db	998	GCATTGAGCTGGACAGAGGC-TACCTCTCTGGGTGACCTCCCTTGTATGGAGCTCTG	940
Qy	357	TTTCATGCTATTTCCTCTATTATAATCTTGCACATG--CACTCTTGTGTGCATGTTTCTT	414
Db	939	T-----TTTCACTCTATTAAATCTTGGAACTGCACACTTTCCTGTGGTCCGTGTGTGT	888
Qy	415	ACGGCTGAGCTGAGCTTTTGTCTCAACGTCACACACTGCTGTTGGCACCAGCCAGACC	474
Db	887	AGGGCTGAGCTGAGCTTTTGTCTGCTGTGCTGCACCACTGCTGTGTTGCTGCCAGACC	828
Qy	475	TGCCGCTGACTCCCATCCCTCTGGAT	500
Db	827	CGCAGCTGACTTTCACCCCTCCGAT	802
RESULT 15			
AAK87491			
ID	AAK87491	standard; DNA; 2629 BP.	
XX	AAK87491;		
AC			
DT	07-NOV-2001	(first entry)	
DE			
XX		Human immune/haematopoietic antigen genomic sequence SRQ ID NO:42303.	
KM		Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KM		cytostatic; gene therapy; vaccine; metastasis; ds.	
XX			
OS	Homo sapiens.		
PN	W0200157182-A2.		
PD			
XX			
XX	09-AUG-2001.		
PF			
XX	17-JAN-2001; 2001WO-US01354.		
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		

PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230457.
PR 06-SEP-2000; 2000US-0230458.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI, 2001-483426/52.
XX
DR
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
XX Disclosure; SEQ ID NO 42303; 3071pp + Sequence listing; English.
XX
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC activity, and can be used in gene therapy and vaccine production. (I)
CC treatment and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 2629 BP; 763 A; 616 C; 587 G; 663 T; 0 other;

Query Match 57.5%; Score 287.6; DB 22; Length 2629;
Best local similarity 84.5%; Pred. No. 8,3e-90;
Matches 377; Conservative 0; Mismatches 49; Indels 20; Gaps 4;

QY 66 ACTGAGAGACAGACTACTGATTTCTTACGCCGACTAAGATTCCTTAAGCTTAGCTGG 125
DB 81 ACTGAGAGACAGACTACTGATTTCTTACGCCGCTTAAGATTCCTTAAGCTTAGCTGG 140
QY 126 GAAGTGACACGCTCCACTTTAAACACGGGCTTGCAACTTAGCTCACACTGACCAAT 185

```

Db 141 GAAATTGACCACTGCACTTTAAACAGGGGCTTGCAATTAGCTCACACCCGACCAAT 200
QY 186 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGGTAAGAAATA 236
Db 201 CAGGTAGTAAAGAGAGCTCACTAAATGCTAATTAGGAAACAGAGGTAAGAAATA 260
QY 237 GCCAATCATCTATTGCTGAGAGCAGACAGAGGGAACAACATCGGATATAAACCCAG 296
Db 261 GCCAATCATCTATTGCTGAGAGCAGACAGAGGGAACAATGATCAGGATATAAACCCAG 320
QY 297 GCATTGAGCTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATGGAGCTGTT 356
Db 321 GCATTGAGCTGGCAACAGACAGCCCTTTGGGTCCCTTCCCTTTGATGGAGCTGTT 379
QY 357 TTCATGCTATTTCACCTGATTAATCTTGAACG--CACTCTGCTGTCAGTTTCTT 414
Db 380 T-----TTCACTCTAATTAATCTTGAACGACACTTTCTGTAGTGTC 431
QY 415 ACGGCTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTGTGGCACACCGAGACC 474
Db 432 ACGGCTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTGTGGCACACCGAGACC 491
QY 475 TGGGCTGACTCCCATCCCTCTGAT 500
Db 492 CACAGCTGACTTCATCCCTCTGAT 517

```

Search completed: April 19, 2003, 12:17:39
 Job time : 197.76 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:03:58 ; Search time 34.6205 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500
Perfect score: 500
Sequence: 1 cccctggggggggcctccctt.....tgacctccatccctctgat 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15333831 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfillseq.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284.8	57.0	2946	4 US-09-175-928-3	Sequence 3, Appl1
2	200	40.0	279	1 US-08-686-878A-50	Sequence 50, Appl1
3	200	40.0	279	1 US-08-721-489-4	Sequence 4, Appl1
4	44	8.8	7218	1 US-08-232-463-14	Sequence 14, Appl1
5	35.4	7.1	7218	1 US-08-232-463-14	Sequence 14, Appl1
6	30.8	6.2	9763	3 US-08-973-273-1	Sequence 1, Appl1
7	30.6	6.1	1078	2 US-08-933-750C-90	Sequence 90, Appl1
8	30.6	6.1	1078	3 US-09-234-613-90	Sequence 90, Appl1
9	30.6	6.1	11827	4 US-09-739-455-3	Sequence 3, Appl1
10	29.8	6.0	16389	4 US-09-741-154-3	Sequence 3, Appl1
11	29	5.8	2023	3 US-08-961-083-199	Sequence 199, App
12	29	5.8	32768	4 US-08-961-527-71	Sequence 71, Appl1
13	29	5.8	36651	4 US-09-738-894A-3	Sequence 3, Appl1
14	28.8	5.8	3386	1 US-08-703-809-2	Sequence 2, Appl1
15	28.8	5.8	3386	1 US-08-703-808-2	Sequence 2, Appl1
16	28.8	5.8	3386	2 US-08-914-066-2	Sequence 2, Appl1
17	28.8	5.8	3386	2 US-08-703-807-2	Sequence 2, Appl1
18	28.8	5.8	3386	2 US-08-747-108A-2	Sequence 2, Appl1
19	28.8	5.8	3386	3 US-09-211-631-2	Sequence 2, Appl1
20	28.8	5.8	3386	3 US-09-265-628-2	Sequence 2, Appl1
21	28.8	5.8	3386	4 US-09-001-141-2	Sequence 2, Appl1
22	28.8	5.8	3386	4 US-09-532-803-2	Sequence 2, Appl1
23	28.6	5.7	2646	4 US-09-066-046-38	Sequence 558, App
24	28.4	5.7	2097	4 US-09-021-017B-558	Sequence 38, Appl1
25	28.4	5.7	2859	5 PCT-US96-05320A-637	Sequence 637, Appl1
26	28	5.6	1262	4 US-08-989-386-4	Sequence 4, Appl1
27	27.8	5.6	15567	4 US-09-627-376-3	Sequence 3, Appl1

28	27.6	5.5	974	1 US-08-848-932-3	Sequence 3, Appl1
29	27.6	5.5	974	2 US-09-008-180-3	Sequence 3, Appl1
30	27.6	5.5	1017	4 US-09-330-611-5	Sequence 5, Appl1
31	27.6	5.5	1666	1 US-08-848-932-1	Sequence 1, Appl1
32	27.6	5.5	1666	2 US-09-008-180-1	Sequence 1, Appl1
33	27.6	5.5	2886	2 US-08-687-080-55	Sequence 55, Appl1
34	27.4	5.5	1631	1 US-08-468-853-5	Sequence 5, Appl1
35	27.4	5.5	1631	1 US-08-468-853-5	Sequence 5, Appl1
36	27.4	5.5	1631	1 US-08-110-357-5	Sequence 5, Appl1
37	27.4	5.5	1631	1 US-08-468-852-5	Sequence 5, Appl1
38	27.4	5.5	1631	2 US-08-468-857-5	Sequence 5, Appl1
39	27.4	5.5	1741	4 US-09-232-160-4	Sequence 4, Appl1
40	27.4	5.5	3054	4 US-09-484-970B-138	Sequence 138, App
41	27.4	5.5	5761	1 US-07-749-001-2	Sequence 2, Appl1
42	27.4	5.5	5761	1 US-08-154-198-2	Sequence 2, Appl1
43	27.4	5.5	5761	1 US-08-463-335-2	Sequence 2, Appl1
44	27.4	5.5	5761	2 US-08-464-023A-2	Sequence 2, Appl1
45	27.2	5.4	360	4 US-09-171-945-54	Sequence 54, Appl1

ALIGNMENTS

RESULT 1
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaValle, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B A1172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-175-928-3
Query Match 57.0%; Score 284.8; DB 4; Length 2946;
Best Local Similarity 95.5%; Pred. No. 5.1e-89;
Matches 315; Conservative 0; Mismatches 12; Indels 3; Gaps 2;
QY 66 ACTGAGACAGAGACTAGTGAATTCCTAGGCGAGTAAGAAATCCCTAAGCTAGCTGG 135
Db 2601 ACTGAGAGACAGAGACTAGTGAATTCCTAGGCTAGCTAAGAAATCCCTAAGCTAGCTGG 2660
QY 126 GAAGGTACCAACGTCACCTTTAAACACGGGGCTTGCAATTGCTACACCTGACCAAT 165
Db 2661 GAAGGTACCAACATCCACCTTTAAACACGGGGCTTGCAATTGCTACACCTGACCAAT 2720
QY 186 CAGAGAGCTCACTAAATATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCAT 245
Db 2721 CAGAGAGCTCACTAAATATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCAT 2780
QY 246 CTATTGCTTAGAGACACAGAGAGGAGCAACATCGGATATTAACCCAGGATTCGAG 305
Db 2781 CTATTGCTTAGAGACACAGAGAGGAGCAACATCGGATATTAACCCAGGATTCGAG 2840
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTTTGTATGGAG--CTGTTTATATGC 363
Db 2841 CTGGCAACAGCA--ACCCCTTTGGGTCCCTTTTGTATGGAGCTGTTTATATGC 2899

Qy 364 TATTCTATTAATCTTGCAACTGCA 393
Db 2900 TATTCTATTAATCTTGCAACTGCA 2929

RESULT 2

US-08-686-878A-50
Sequence 50, Application US/08686878A
Patent No. 5708157
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavalley, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-686-878A-50

Query Match 40.0%; Score 200; DB 1; Length 279;

Best Local Similarity 88.2%; Pred. No. 4.8e-60;
Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;

Qy 133 ACCAGTCCACCTTTAAACAGGGGCTTGCACTAGC^CACACCGACCAATCAGAGAG 192
Db 1 RCCACATCCACCTTTAAACAGGGGCTTGCAANAAAG^NACACTTGACCAATCAGAGAG 60
Qy 193 CTCCTAAATATGCTAATTAGGCAAGAGAGGTAAC^AAATAGCAATCATCTATTGC 252
Db 61 NTCANTAAATATATATTNMGCAAAAACGAGGTAAG^AAATAGCAATCATCTATTGC 120
Qy 253 CTGAGAGCAGCAGAGAGGGAACAATGGGATATAA^CCAGGCAATTCAGCTGGCAA 312
Db 121 CTGAGAGCAGCAGAGAGGGAACAATGATCGGATATAA^CCAGGTTTNGAGCCGGCAA 180
Qy 313 CAGCAGCCCCCTTTGGGCTCCCTTCTGTATAGGAGC^--TGTTTCATGCTATTTC 370
Db 181 CGGCA-ACCCCTTTGGGCTCCCTTCTGTATAGGAGC^TNTGTTTCATGCTATTTC 239
Qy 371 CTCTATTAAATCTTGCAACTGCA 393

Db 240 NTNTATTAAATNTTGCAACTGCA 262

RESULT 3

US-08-721-489-4
Sequence 4, Application US/08721489
Patent No. 5786465
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavalley, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-721-489-4

Query Match 40.0%; Score 200; DB 1; Length 279;

Best Local Similarity 88.2%; Pred. No. 4.8e-60;
Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;

Qy 133 ACCAGTCCACCTTTAAACAGGGGCTTGCACTAGC^CACACCGACCAATCAGAGAG 192
Db 1 RCCACATCCACCTTTAAACAGGGGCTTGCAANAAAG^NACACTTGACCAATCAGAGAG 60
Qy 193 CTCCTAAATATGCTAATTAGGCAAGAGAGGTAAC^AAATAGCAATCATCTATTGC 252
Db 61 NTCANTAAATATATATTNMGCAAAAACGAGGTAAG^AAATAGCAATCATCTATTGC 120
Qy 253 CTGAGAGCAGCAGAGAGGGAACAATGGGATATAA^CCAGGCAATTCAGCTGGCAA 312
Db 121 CTGAGAGCAGCAGAGAGGGAACAATGATCGGATATAA^CCAGGTTTNGAGCCGGCAA 180
Qy 313 CAGCAGCCCCCTTTGGGCTCCCTTCTGTATAGGAGC^--TGTTTCATGCTATTTC 370
Db 181 CGGCA-ACCCCTTTGGGCTCCCTTCTGTATAGGAGC^TNTGTTTCATGCTATTTC 239
Qy 371 CTCTATTAAATCTTGCAACTGCA 393
Db 240 NTNTATTAAATNTTGCAACTGCA 262

Db 1055 CCT 1053

RESULT 6
US-08-973-273-1/c

; Sequence 1, Application US/08973273
; Patent No. 6140085

; GENERAL INFORMATION:

; APPLICANT: Dean, Caroline

; APPLICANT: MacKnight, Richard C

; APPLICANT: Bancroft, Ian

; APPLICANT: Lister, Clare K

; TITLE OF INVENTION: Genetic Control of Flowering

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Nixon & Vanderhye P.C.

; STREET: 1100 No. 6140085th Glebe Road, 8th Floor

; CITY: Arlington

; STATE: Virginia

; COUNTRY: United States of America

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/973,273

; FILING DATE: 01-DEC-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB96/01332

; FILING DATE: 03-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 951196.9

; FILING DATE: 02-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Ms Mary J Wilson

; REGISTRATION NUMBER: 32,955

; REFERENCE/DOCKET NUMBER: 620-29

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 816-4000

; TELEFAX: (703) 816-4100

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 9763 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; HYPOTHEetical: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Arabidopsis thaliana

; STRAIN: Columbia

; US-08-973-273-1

Query Match 6.2%; Score 30.8; DB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 3.5;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 176 CCTGACCAATGAGAGCCACTAAATGCTAATAGCAGAGAGAGGTAAGAAT 235
DB 4845 CTTGAGAGAGAGAGAACTATATCTGATGTAAGCAGAGAGAGAGAGAGAA 4786

QY 236 AGCCAATCATCTAT 249
DB 4785 GGATCAATCATCAT 4772

RESULT 7
US-08-933-750C-90
; Sequence 90, Application US/08933750C

; Patent No. 5932442

; GENERAL INFORMATION:

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Bandman, Olga

; APPLICANT: Shah, Purni

; APPLICANT: Au-Young, Janice

; APPLICANT: Yue, Henry

; APPLICANT: Guegler, Karl J.

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES

; NUMBER OF SEQUENCES: 98

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/933,750C

; FILING DATE: September 23, 1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0356 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 90:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1078 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: SINTNOT01

; CLONE: 2184712

; US-08-933-750C-90

Query Match 6.1%; Score 30.6; DB 2; Length 1078;
Best Local Similarity 58.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGGCAATCGAGCTGCAACAGCAGCCCGCTTGGGTCCTTCCCTTGTATG 347
DB 941 TCACTCGAAGAGAGAACTCTGTAAGAGAGAGCCCTTGGGTCCTTCTTCTT---TG 996

QY 348 GAGCTGTTTCAATGCTATTTCACTCTATTTAAATCTTGCAATGCACTCTCTGTCAT 407
DB 997 ATAGAGATTATTAATGCTTGTCTTCCCAATAACTGGGCAATGGAATCCATGTCAT 1056

QY 408 GTTCTT 414
DB 1057 ACTGCT 1063

RESULT 8
US-09-234-613-90
; Sequence 90, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti

APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Puryi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 90:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SININOT01
CLONE: 2184712
US-09-234-613-90

Query Match 6.1%; Score 30.6; DB 3; Length 1078;
Best Local Similarity 58.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGCATTGAGCTGGCAAGCAAGCCCTTTGGCTCCTTCCCTTTGATG 347
DB 941 TCAACTGTGAAGAGATCTTGTCAAGAGAGCCCTTGCGCTCCCTTCTT---TG 996
QY 348 GGAAGCTTTTCAATGCTATTTCACTATTAATCTTGAACCTGACCTCTTGGTCCAT 407
DB 997 ATAGCAATTATATGCTTCTTGTCCCAATAAATGCGGAGATGAAATCTAGTCTAT 1056
QY 408 GTTTCCT 414
DB 1057 ACTGCCT 1063

RESULT 9
US-09-739-455-3
Sequence 3, Application US/09739455
Patent No. 6413756
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL000653
CURRENT APPLICATION NUMBER: US/09/739,455
CURRENT FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 11827
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(11827)
OTHER INFORMATION: n = A,T,C or G
US-09-739-455-3

Query Match 6.1%; Score 30.6; DB 4; Length 11827;
Best Local Similarity 51.1%; Pred. No. 4.6;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 117 CCTAGCTGGAGAGTGACACGCTCCACTTTTAAACAGGGGCTTGCACTTACCTACAC 176
DB 10853 CCGAGCTGCTGGAGTGGAAGTCTGCTTTGTTGGCGGCTTGTCTTAATCAGTT 10912
QY 177 CTGACCAATCAGAGAGCTCACTAAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATA 236
DB 10913 CCTCTTAGATTATTATACACTAAAAAAATTAAGTTTGAAGAAATAGAGATA 10972
QY 237 GCCAATCATCTATTCCTGANG 257
DB 10973 CAGAAACATGAATTTCACGAG 10993

RESULT 10
US-09-741-154-3/C
Sequence 3, Application US/09741154
Patent No. 6437110
GENERAL INFORMATION:
APPLICANT: BEASLEY, Ellen M. et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001061
CURRENT APPLICATION NUMBER: US/09/741,154
CURRENT FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 16389
TYPE: DNA
ORGANISM: Human
US-09-741-154-3

Query Match 6.0%; Score 29.8; DB 4; Length 16389;
Best Local Similarity 63.0%; Pred. No. 11;
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 74 ACAGAGCTAGCTGATTTCTTAGCCGACCTAAGATCCCTAAGCTTAGGGAAGGTGA 133
DB 1196 ACAGACTAATGTGAATGCTGAGTCAAAATTAATTCCTGATCCAGCTGGGAAGGG 1137
QY 134 CCAGTCCACCTT 146
DB 1136 ATGCTCGAGCTT 1124

RESULT 11
US-08-961-083-199
Sequence 199, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452

```

CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 2023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-199

Query Match
Best Local Similarity 5.8%; Score 29; D3 3; Length 2023;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTATTAGCAAGACAGAGGTAAGAATAGCCATCATCTATTGCTGAGAGCAC 262
DB 1711 TGATTTCACTTTAAGATAGACGCTAGAGTATAGTACTGATTCATTGGCTCAGATGAC 1770

QY 263 A 263
DB 1771 A 1771

RESULT 12
US-08-961-527-71
Sequence 71, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
```

```

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-71

Query Match
Best Local Similarity 5.8%; Score 29; DB 4; Length 32768;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTATTAGCAAGACAGAGGTAAGAATAGCCATCATCTATTGCTGAGAGCAC 262
DB 7947 TGATTTCACTTTAAGATAGACGCTAGAGTATAGTATGATGATGCTGCTCAGATGAC 8006

QY 263 A 263
DB 8007 A 8007

RESULT 13
US-09-738-894A-3/C
Sequence 3, Application US/09738894A
Patent No. 6331423
GENERAL INFORMATION:
APPLICANT: GUGGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL000636
CURRENT APPLICATION NUMBER: US/09/738,894A
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PASCSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 36651
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(36651)
OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match
Best Local Similarity 5.8%; Score 29; DB 4; Length 36651;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 15 TCCTTTCTGGATGAGGCAAAAGCCCTGAGATACAGCAATTATCTTGCACATGAGAGA 74
DB 6891 TCCTACTCGAACAATGAGCCCTGATGCTGGGGCTACAGCAACCATCTTGCAGCATGGA 6832

QY 75 CAGGACTAG 83
DB 6831 AAAGGCTAG 6823

RESULT 14
US-08-703-809-2
Sequence 2, Application US/08703809
Patent No. 5716808
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: GENETIC ENGINEERING OF PICHIA METHANOLICA
```

```

; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,809
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 96-18
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-703-809-2

```

```

Query Match      5.8%; Score 28.8; DB 1; Length 3386;
Best Local Similarity 49.3%; Pred. No. 9.6;
Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 307 TGGCAACAGCAGCCCCCTTGGGTCCTTCCTTTGTATGGAGCTGTTTCATGCTAT 366
    |||||
DB 856 TGGCAACAGGATAGTCTATCTGCTTATTCATCCACTTGGGAACGCTCTCTTAC 915
    |||||
QY 367 TTCACTATTAATCTTGAACGCACTGCTTCTGTCATGTTTCTAGCGCTCGAGCT 426
    |||||
DB 916 CCCAGATTCTCAAGCTAATATCTGCCCTTGTCTATTTGCTTTCCTGTAACAAGCG 975
    |||||
QY 427 GAGCTTTGCTCAACGTCACCACTGCTGTT 458
    |||||
DB 976 GAGCTTTGCTCCCATCTCTTGTGTTT 1007
    |||||

```

```

RESULT 15
US-08-703-808-2
; Sequence 2, Application US/08703808
; Patent No. 5736383
; GENERAL INFORMATION:
; APPLICANT: Raymond, Christopher K.
; TITLE OF INVENTION: PREPARATION OF PICHIA METHANOLICA AUXOTROPHIC
; TITLE OF INVENTION: MUTANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,808
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 96-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-703-808-2

```

```

Query Match      5.8%; Score 28.8; DB 1; Length 3386;
Best Local Similarity 49.3%; Pred. No. 9.6;
Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 307 TGGCAACAGCAGCCCCCTTGGGTCCTTCCTTTGTATGGAGCTGTTTCATGCTAT 366
    |||||
DB 856 TGGCAACAGGATAGTCTATCTGCTTATTCATCCACTTGGGAACGCTCTCTTAC 915
    |||||
QY 367 TTCACTATTAATCTTGAACGCACTGCTTCTGTCATGTTTCTAGCGCTCGAGCT 426
    |||||
DB 916 CCCAGATTCTCAAGCTAATATCTGCCCTTGTCTATTTGCTTTCCTGTAACAAGCG 975
    |||||
QY 427 GAGCTTTGCTCAACGTCACCACTGCTGTT 458
    |||||
DB 976 GAGCTTTGCTCCCATCTCTTGTGTTT 1007
    |||||

```

Search completed: April 19, 2003, 14:48:55
Job time : 68.6205 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 19, 2003, 12:08:18 ; Search time 75.5659 Seconds
(without alignments)
6646.925 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500
Sequence: 1 cccctgggggggctctctt.....tgacctccatccctcgat 500

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	348.2	69.6	1329	7	US-08-979-847-108
2	301	60.2	180557	12	US-10-003-806-6
3	301	60.2	180557	12	US-10-003-806-9
4	284.8	57.0	2946	9	US-10-114-893-134
5	283.8	56.8	2930	10	US-09-902-535-1
6	270.4	54.1	635	7	US-08-979-847-102
7	261.4	52.3	410	10	US-09-880-107-538
8	259.2	51.8	541	10	US-09-864-761-8173
9	232.2	46.4	326014	10	US-09-731-231A-3
10	226.2	45.2	1894	10	US-09-864-761-4444
11	218	43.6	15425	9	US-10-091-504-1654
12	218	43.6	15425	9	US-09-764-869-1654
13	204.6	40.0	569	10	US-09-864-761-14951
14	200	40.0	279	12	US-10-040-916-50
15	126	25.2	246	10	US-09-864-761-20462
16	118.2	23.6	446	10	US-09-811-284-3
17	111.6	22.3	475	10	US-09-864-761-895
18	109.4	21.9	3524	10	US-09-972-724-1
19	102	20.4	409	10	US-09-864-761-4153

C 20	101	20.2	559	10	US-09-864-761-7501	Sequence 7501, Ap
C 21	97.8	19.6	504	10	US-09-864-761-7027	Sequence 7027, Ap
C 22	93	18.6	440	10	US-09-864-761-3694	Sequence 3694, Ap
C 23	79.2	15.8	579	10	US-09-864-761-13678	Sequence 13678, A
C 24	50.4	10.1	88	10	US-09-864-761-20907	Sequence 20907, A
C 25	49.8	10.0	525	10	US-09-893-737-31	Sequence 31, Appl
C 26	48.8	9.8	387	10	US-09-864-761-30194	Sequence 30194, A
C 27	37.8	7.6	275	10	US-09-864-761-17675	Sequence 17675, A
C 28	34.6	7.0	1568	9	US-09-735-713A-7	Sequence 7, Appl1
C 29	34.6	6.9	79	10	US-09-864-761-31488	Sequence 31488, A
C 30	32.6	6.5	384	10	US-09-783-590-10544	Sequence 10544, A
C 31	32.6	6.5	653	9	US-10-184-644-402	Sequence 402, App
C 32	32.6	6.5	653	9	US-10-184-634-402	Sequence 402, App
C 33	31.8	6.4	468	10	US-09-796-692-4666	Sequence 4666, Ap
C 34	31.2	6.2	473	10	US-09-864-761-11001	Sequence 11001, A
C 35	31.2	6.2	10514	10	US-09-764-877-3470	Sequence 3470, Ap
C 36	31.2	6.2	197997	10	US-09-822-246-3	Sequence 3, Appl1
C 37	31	6.2	407	10	US-09-864-761-20523	Sequence 20523, A
C 38	30.8	6.2	2000	9	US-09-938-842A-2796	Sequence 2796, Ap
C 39	30.6	6.1	410	10	US-09-867-701-4129	Sequence 4129, Ap
C 40	30.6	6.1	1078	10	US-09-840-787-90	Sequence 90, Appl
C 41	30.6	6.1	8220	10	US-09-797-908-3	Sequence 3, Appl1
C 42	30.6	6.1	8522	10	US-09-817-181-3	Sequence 3, Appl1
C 43	30.2	6.0	414	9	US-10-123-153-418	Sequence 418, App
C 44	30.2	6.0	2012	9	US-09-746-783-147	Sequence 147, App
C 45	30.2	6.0	2014	9	US-10-036-041-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-979-847-108
; Sequence 108, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUKE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OLIFF & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 108:

QY	266	AGGAGGGAACAACAAATCGGAAATTTAAACCCAGGCAATTCGAGCTGGGCAACGAGCCCCCT	325
Db	58782	AGGAGGGAACAATGATTCAGATTTAAACCCAGGCAATTCGAGCAACGAGCAGC-TACGCTTT	58724
QY	326	TTGGTCCCTTCCTCTTGTATGGAGAGCTGTTTCATGCTATTTCACCTATTAAATCTTG	385
Db	58723	TTGGTCCCTTCCTCTTGTATGGAGAGCTGTG-----CTTCACTATTAAATCTTG	58672
QY	386	CAACTGCA--CTCTTCTGTGCATGTTTCTACGGCTGAGCTGAGCTTTTGCTACCGT	443
Db	58671	CAGCTGCACTCTCTTTTGATTCACATTTGTCAATGTTTCAGAGCTGAGCTTCTCTGCCGT	58612
QY	444	CCACCACTGCTGTTTGCCACCAACCGAGACCTGCCGTGACTCCCATCCCTCGAT	500
Db	58611	CCACCACTGCTGTTTGCCGCTGTCCGACACTGCTGCTGACTTCCATCCGTCCAGAT	58555

```

RESULT 4
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKenough
; APPLICANT: Kelleher, Kerry S.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; CURRENT FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; EARLIER FILING DATE: 1999-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-134

```

Query Match	57.0%;	Score 284.8;	DB 9;	Length 2946;
Best Local Similarity	95.5%;	Pred. No. 2.6e-89;		
Matches 315;	Conservative	0;	Mismatches 12;	Indels 3;
				Gaps 2;

Oy	66	ACTGAGAGACAGGACTGAGCTGGAAATTTCTTAGGCGGACTGAAGATCCCTAAGCTGAGCTGG	125
Db	2661	ACTGAGAGACAGGACTGAGCTGGAAATTTCTTAGGCTGAAGATCCCTAAGCTGAGCTGG	2660
Oy	126	GAAGGTACCACTGTCACCTTTAAACAAGGGGCTTGCAACTTAGCTCACACTGACCAAT	185
Db	2661	GAAGGTACCACTGTCACCTTTAAACAAGGGGCTTGCAACTTAGCTCACACTGACCAAT	2720
Oy	186	CAGAGAGCTCACTAAATGTCTAATTAGGCAAAACAGAGAGGTAAAGAAATAGCCAAAT	245
Db	2721	CAGAGAGCTCACTAAATGTCTAATTAGGCAAAACAGAGGTAAAGAAATAGCCAAAT	2780
Oy	246	CTATTGCTGTGAGCAGACAGAGAGGGACAACTGGGATATAAACCCAGGCAATTGAG	305
Db	2781	CTATTGCTGTGAGCAGACAGAGAGGGACAACTATGGGATATAAACCCAAATCTTTCAG	2840
Oy	306	CTGGCAACAGAGGCCCCCTTTGGGATCCCTTCCTTTGTATGGGAG--CTGTATTCATGC	363
Db	2841	CCGGCAACAGGCA-ACCCCTTTGGGATCCCTTCCTTTGTATGGGAGCTGTGTTTTCATGC	2895
Oy	364	TATTTCACCTATTAAATCTTGCACTGCA	393

Db 2900 TATTCTCTATTAACTTGCACCTGCA 2929

```

RESULT 5
US-09-902-535-1
; Sequence 1, Application US/09902535
; Patent No. US20020102530A1
; GENERAL INFORMATION:
; APPLICANT: Keith, Jr., James C.
; APPLICANT: McCoy, John M.
; APPLICANT: Mt, Sna
; TITLE OF INVENTION: Methods and compositions for diagnosing
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
; TITLE OF INVENTION: disorders
; FILE REFERENCE: GIN-6006B4
; CURRENT APPLICATION NUMBER: US/09/902,535
; CURRENT FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,657
; PRIOR FILING DATE: 2000-07-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (930)...(2546)
US-09-902-535-1

```

Query Match	56.8%;	Score 283.8;	DB 10;	Length 2930;
Best Local Similarity	-95.4%;	Pred. No. 5.8e-89;		
Matches 314; Conservative	0;	Mismatches 12;	Indels 3;	Gaps 2;

Qy	66	ACTAGAGACAGGACTAGCTGAGATTCTCTAGGCGCACTAGAAATCCCTAAGCTGCTGG	125
Qy	2603	ACTGAGGACAGAGCTTAGCTGAGTTTCTTAGGCTGACTAGAAATCCCTAAGCTGCTGCG	2662
Qy	126	GAGGTAACCAACGCTCACTTTTAAACAGCGGGCTTGCACTTAGCTCACACTGACCAAT	185
Db	2663	GAGGTACCAACTCCACTTTTAAACAGCGGGCTTGCACTTAGCTCACACTGACCAAT	2722
Qy	186	CAGAGACTCACTAAATGCTAATTAGGCAAAACAGAGGTTAAAGAAATAGCCAACTAT	245
Db	2723	CAGAGACTCACTAAATGCTAATTAGGCAAAAACAGAGGTTAAAGAAATAGCCAACTAT	2782
Qy	246	CTATTGCTTAGAGACACAGCAGAGGGGACAACTATGGGATTTAAACCGAGCACTTGAG	305
Db	2783	CTATTGCTTAGAGACACAGCAGAGGGGACAACTATGGGATTTAAACCGAGCTTTGAG	2842
Qy	306	CTGCGCACACAGACCCCCCTTTGGGTCCCTTCCCTTTGATGGGAG--CTGTTTTCAATGC	363
Db	2843	CCGCGCAACGGCA--ACCCCTTTGGGTCCCTTCCCTTTGATGGGAGCTCTGTTTTCAATGC	2901
Qy	364	TATTTCACCTATTAAATCTTGCACTGC	392
Db	2902	TATTTCACCTATTAAATCTTGCACTGC	2930

```

RESULT 5
US-08-979-847-102
, Sequence 102, Application US/08979847
, Publication No. US2003039664A1
, GENERAL INFORMATION:
, APPLICANT: PERRON, HEVRE
, APPLICANT: BESEME, FREDERIC
, APPLICANT: BEDIN, FREDERIC
, APPLICANT: PARANHOS-BACCALA, GLAUCIA
, APPLICANT: KOMRIN-PADELI, FLORENCE
, APPLICANT: JOLIVET-REYNAUD, COLETTE
, APPLICANT: GARRON, BERNARD
, APPLICANT: GARRON, JEREMY

```

APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIVE & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-102

Query Match 54.1%; Score 270.4; DB 7; Length 635;
Best Local Similarity 92.7%; Pred. No. 1.3e-84;
Matches 306; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

QY 66 ACTGAGACAGAGCTAGCTGATTTCTTAGGCCCACTAAGATCCCTAGCTGCTGG 125
DB 231 ACTGAGACAGAGCTAGCTGATTTCTTAGGCCCACTAAGATCCCTAGCTGCTGG 350
QY 126 GAAGGTGACAGCTGACCTTTAAACAGGGGCTGCACTAGCTCAGACCTGCAAT 185
DB 351 GAAGGTGACAGCTGACCTTTAAACAGGGGCTGCACTAGCTCAGACCTGCAAT 410
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCMAAGACAGAGGTAAAGAAATAGCAATCAT 245
DB 411 CAGAGAGCTCACTAAATGCTAATTAGGCMAAGAAATAGAGTAAAGAAATAGCAATCAT 470
QY 246 CTATTGCTGAGAGCAGCAGAGGAGGACAAATCGGATTAATTAACCCAGGCAATTCGAG 305
DB 471 CTATTGCTGAGAGCAGCAGGAGGAGGACAAAGATCGGATTAATTAACCCAGGCAATTCGAG 530
QY 306 CTGGCAACAGAGCCCTTTGGGCTCCCTTGTATGGGAG--CTGTTTCATGCG 363
DB 531 CTGGCAACAGAGC--ACCCCTTTGGGCTCCCTTGTATGGGAGCTCTGTTTCACCT 589
QY 364 TATTTCACCTATTAATCTTGCAACTGCA 393
DB 590 TATTTCACCTATTAATCTTGCAACTGCA 619

RESULT 7
US-09-880-107-538/c
Sequence 538, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darcil T.
APPLICANT: Vockley, Joseph G.

APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 538
LENGTH: 410
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538

Query Match 52.3%; Score 261.4; DB 10; Length 410;
Best Local Similarity 91.2%; Pred. No. 1.4e-81;
Matches 300; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

QY 66 ACTGAGACAGAGCTAGCTGATTTCTTAGGCCGACTAAGATCCCTAGCTGCTGG 125
DB 329 ACTGAGACAGAGCTAGCTGATTTCTTAGGCCGACTAAGATCCCTAGCTGCTGG 270
QY 126 GAAGGTGACAGCTGACCTTTAAACAGGGGCTGCACTAGCTCAGACCTGCAAT 185
DB 269 GAAGGTGACAGCTGACCTTTAAACAGGGGCTGCACTAGCTCAGACCTGCAAT 210
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCMAAGACAGAGGTAAAGAAATAGCAATCAT 245
DB 209 CAGAGAGCTCACTAAATGCTAATTAGGCMAAGAAATAGAGTAAAGAAATAGCAATCAT 150
QY 246 CTATTGCTGAGAGCAGCAGAGGAGGACAAATCGGATTAATTAACCCAGGCAATTCGA 304
DB 149 CTATTGCTGAGAGCAGCAGGAGGAGGACAAAGATTCGAAATTAATTAACCCAGGCAATTCGA 90
QY 305 GCTGGCAACAGAGCCCTTTGGGCTCCCTTGTATGGGAG--CTGTTTCATGCG 362
DB 89 GCTGGCAACAGAGCCCTTTGGGCTCCCTTGTATGGGAGCTCTGTTTCACCT 30
QY 363 CTATTTCACCTATTAATCTTGCAACTG 391
DB 29 CTATTTCACCTATTAATCTTGCAACTG 1

RESULT 8
US-09-864-761-8173/c
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27


```

1 PRIOR APPLICATION NUMBER: PCT/US01/00666
2 PRIOR FILING DATE: 2001-01-30
3 PRIOR APPLICATION NUMBER: PCT/US01/00667
4 PRIOR FILING DATE: 2001-01-30
5 PRIOR APPLICATION NUMBER: PCT/US01/00664
6 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00669
8 PRIOR FILING DATE: 2001-01-30
9 PRIOR APPLICATION NUMBER: PCT/US01/00665
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00668
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: PCT/US01/00663
14 PRIOR FILING DATE: 2001-01-30
15 PRIOR APPLICATION NUMBER: PCT/US01/00662
16 PRIOR FILING DATE: 2001-01-30
17 PRIOR APPLICATION NUMBER: PCT/US01/00661
18 PRIOR FILING DATE: 2001-01-30
19 PRIOR APPLICATION NUMBER: PCT/US01/00670
20 PRIOR FILING DATE: 2001-01-30
21 PRIOR APPLICATION NUMBER: US 60/234,687
22 PRIOR FILING DATE: 2000-09-21
23 PRIOR APPLICATION NUMBER: US 09/608,408
24 PRIOR FILING DATE: 2000-06-30
25 PRIOR APPLICATION NUMBER: US 09/774,203
26 PRIOR FILING DATE: 2001-01-29
27 NUMBER OF SEQ ID NOS: 49117
28 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
29 SEQ ID NO 8173
30 LENGTH: 541
31 TYPE: DNA
32 ORGANISM: Homo sapiens
33 FEATURE:
34 OTHER INFORMATION: MAP TO AC016663.2
35 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
36 OTHER INFORMATION: EXPRESSED IN HEILA, SIGNAL = 3.8
37 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
38 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
39 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
40 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
41 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
42 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
43 US-09-864-761-8173
44
45 Query Match 51.8%; Score 259.2; DB 10; Length 541;
46 Best Local Similarity 81.6%; Pred. No. 1e-80;
47 Matches 367; Conservative 0; Mismatches 63; Indels 20; Gaps
48
49 QY 62 TGCACTGAGAGACAGGACTAGCTGATTTCTTAGCCGACCTAAGAAATCCCTAACCTTAG 121
50 DB 537 TGGCGTAGAGACAGGAGCTAGCTGATTTCTTAGCCGACCTAAGAAATCCCTAACCTTAG 478
51 QY 122 CTGGGAAGGTGACACGATCCACTTTAAACAGGGGGCTTGCAACTTAGCTCACACTGAC 181
52 DB 477 CTGGGAAGGTGACACCACTCTACCTTGAACACAGGGCTTGCAACTTGGCTCACACTGAC 418
53 QY 182 CAATC-----AGAGAGCTCACTAAATGCTTAATTAGGCAAAACAGAGGTTAAGA 232
54 DB 417 CAATCAGCTAGTAAAGACAGCTCACTAAAGGCTAATTCGGCTAAACAGAGGTTAATA 358
55 QY 233 AATTGCAATATCATCTATGCTCCGTAAGACACAGAGAGGAGAAACAATGGGATTAAC 292
56 DB 357 AATTGCAATATCATCTATCCTGTAAGACAGAGGGAGTGAATTAATGATCGGATCAAC 298
57 QY 293 CCAGGCAATTGAGCTGGCAACAGCAGACCCCTTTGGGCTCTTCCGCTTTGATGGAGAC 352
58 DB 297 CCAGGCAATTGAGCTGGCAACAGCAGATGCA-ACGCCCTTAGGGTCCCTCCCACTGATGGAGC 239
59 QY 353 TGTATTCAATGCTATTCTCTATTAAATCTTGAACCTGCACTCT-TCGTGTCATGTTT 411
60 DB 238 TCTGT-----TTCACTCTATTAAATCTTGAACCTGCAACACTCTGTGTGTGTTT 187
61 QY 412 CTTAGCGCTCAGCTGAGCTTTTGTCTACCGGTCCACCACTGCTGTGTTTGACACACCGCAG 471

```

```

Db      186  ATTCCGGTTGCAATAGAGCTTTTCTCCACCATCACCACCTGCTGGAATAGCTGTCTTGAG 127
              |||
QY      472  AC-CTGCCGCTGACTCCCATCCCTCTTGAGT 500
              |||
Db      126  ACACCGCTGCTGACTTCCACCCCTCCGGAT 97
              |||

RESULT 9
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THERBOF
; FILE REFERENCE: CL001007
; CURRENT APPLICATION NUMBER: US/09/731,231A
; CURRENT FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 326014
; TYPE: DNA
; ORGANISM: Human
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (1)...(326014)
; OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match      46.4%; Score 232.2; DB 10; Length 326014;
Best Local Similarity 86.7%; Pred. No. 1.2e-69;
Matches 294; Conservative 0; Mismatches 33; Indels 12; Gaps 3;

QY      66  ACTAGAGACAGAGACTAGCTGGAATTTCTCTAGGCGCACTAAGAAATCCCTAAGCCTACTGG 125
              |||
Db 170532  ACTAGAGACAGAGACTAGCTGGAATTTCTCTAGGCTAAGAAATCCCTAAGCCTACTG 170473
              |||
QY      126  GAAGGTACCAACGTCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAAT 185
              |||
Db 170472  GAAGGTACCTGCTTCTTAAACCCGGGCTTGCACTTAGCTCAGACCTGACCAAT 170413
              |||
QY      186  C-----AGAGGCTCACTAAAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATA 236
              |||
Db 170412  CAGGTAGAAAGAGAGCTCACTAAAAATGCTAATTAGGCTAAGCAAGAGTAAAGAAATA 170353
              |||
QY      237  GCCAATATCTATTGCTGTAGAGACAGACAGAGGAGCAACAATCGGATATTAACCCAG 296
              |||
Db 170352  GCCAATATCTATGCTGTAGAAACAGTGGAGGAGCAATGATTTGGTTATTAACCCAG 170293
              |||
QY      297  GCATTGAGCTGGCAACAGAGAGCCCTTTGGGTCCTTCCCTTTGTATGGAG--CTG 354
              |||
Db 170292  GCATTGAGCTGGCAAGAGGCG-TACCAATTTTGGGTCCTTCCCTTTGTATGGAGCTCTG 170234
              |||
QY      355  TTTTCATGCTATTTCACCTCTATTAAATCTTGCACCTGCA 393
              |||
Db 170233  TTTTCACCTATTTCACCTATTAAATCTTGCACAGCA 170195
              |||

RESULT 10
US-09-864-761-4444
; Sequence 4444, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OR INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Neomica-X-1

```

```

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine ver.8.1.1
SEQ ID NO 4444
LENGTH: 1894
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC002346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444

Query Match      45.2%; Score 226.2; DB 10; Length 1894;
Best Local Similarity 86.4%; Pred. No. 8.2e-69;
Matches 291; Conservative 0; Mismatches 28; Indels 18; Gaps 3;

```

```

237 GCCAATCATCTATTGCTGTAGAGACAGAGAGGAGCAACATCGGATATAAACCAG 236
1687 GCCAATCATCTATTGCTGTAGAGACAGAGAGGAGCAATCGGATATAAACCAG 1746
297 GCATTGAGCTGGCAGACAGAGCCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGT 356
1747 GCATTGAGCTGGCAGACAG-TACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGT 1805
357 TTCAATGCTATTTCACCTCTATTAAATCTTGCACTGCA 393
1806 T-----CTTCACTCTATTAAATCTTGCACTGCA 1834

RESULT 11
US-10-091-504-1654/c
Sequence 1654, Application US/10091504
Publication No. US2003005908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1654

Query Match      43.6%; Score 218; DB 9; Length 15425;
Best Local Similarity 85.5%; Pred. No. 2e-65;
Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

```

NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1654

Query Match 43.6%; Score 218; DB 10; Length 15425;
Best Local Similarity 85.5%; Pred. No. 2e-65;
Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

QY 71 GAGACAGAGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTTAGGCTGAGGAGG 130
DB 1998 GAGACAGAGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTTAGGCTGAGGAGG 1939
QY 131 TGACGACGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTTAGGCTGAGGAGG 186
DB 1938 TGACGACGCTAGTGGATTCTTCTAGGCGCACTAAGATCCCTTAGGCTGAGGAGG 1879
QY 187 -----AGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGTAAAGAAATAGCCAA 241
DB 1878 ATTAAGAGAGCTCAAGAAATGCTAATTAGGCAAGAGAGTAAAGAAATAGCCAA 1819
QY 242 TCATCTATTCTCTAGAGCAGCAGGAGGAGCAACATCGGATATPAAACCCAGGCAAT 301
DB 1818 TCATCTATTCTCTAGAGCAGCAGGAGGAGCAACATCGGATATPAAACCCAGGCAAT 1759
QY 302 CGAGCTGGCAAGCAGCCCTTGGGTCCTTCCCTTGTATGGAGCTGTTTTCAT 361
DB 1758 TGAGCTGGCAAGC-TACCCCTTTTGGGTCCTTCCCTTGTATGGAGCTGTTTTCAT 1704
QY 362 GCTATTTCACTTATTAATCTTGCAACTGCA 393
DB 1703 -----TTTCACTTATTAATCTTGCAACTGCA 1676

RESULT 13
US-09-864-761-14951/c
Sequence 14951, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmlca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIORITY FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14951
LENGTH: 569
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000233.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
US-09-864-761-14951

Query Match 40.9%; Score 204.6; DB 10; Length 569;
Best Local Similarity 79.4%; Pred. No. 1.6e-61;
Matches 336; Conservative 0; Mismatches 64; Indels 23; Gaps 7;

QY 66 ACTGAGAGACAGGACTAGTGGATTCTTCTAGGCGCACTAAGATCCCTTAGGCTGAG 125
DB 415 ACCGAGAGACAGGACTAGTGGATTCTTCTAGGCGCACTAAGATCCCTTAGGCTGAG 356
QY 126 G-AAGGTGACGACGCTTAAACAGGAGGCTTGCACTTAGCTACACCTGACCA 184
DB 355 GAAAGGTGACGACGCTTAAACAGGAGGCTTGCACTTAGCTACGCTCGATCGACCA 296
QY 185 TC-----AGAGAGCTCACTAAATGCTAATTGAGC-AAAGACAGAGGTTAAAGAA 224
DB 295 TCAGGTAGTAAAGAGGCTTCACTGAATTAATTTAGGCTTAAAGAGAGGTTAAAGAA 236
QY 235 TAA-CCAATCATCTATTGCTGAGAGCAGCAGGAGGAGCAACATCGGATAT-AAAC 292
DB 235 TAGTCAAATCATCTATTGCTGAGAGCAGCAGGAGGAGCAATGATTTGGATATTAAC 176
QY 293 CCAGGATTCGAGCTGAGCAAGCAGCAGCCTTGGGTCCTTCCCTTTGTATGGAGC 352
DB 175 CCAAGGATTCGAGCTGAGCAAGCAGCAGCCTTGGGTCCTTCCCTTTGTATGGAGC 116
QY 353 TGTTCAGCTATTCTTCACTTAAATCTTGCAATG--CACTCTTGTGTCATGTT 410
DB 115 TCTGT-----TTTCACTCTGTTAAATCTTGCACTGACCTCTCTGTCAGTGT 64
QY 411 TCTTACGCTGAGCTGAGCTTGTGCTACCCGTCACCACTGCTGTTTGGACACCGCA 470
DB 63 TGTTCGAGCTGAGCTGAGCTTGTGCTACCCGTCACCACTGCTGTTTGGACACCGCA 4
QY 471 GAC 473
DB 3 GAC 1

RESULT 14
US-10-040-916-50
Sequence 50, Application US/10040916
Patent No. US2002014679A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallie, Edward

Racie, Lisa
Meiberg, David
Treacy, Maurice
Evans, Cheryl
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,916
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,029
FILING DATE: 07-FEB-1997
APPLICATION NUMBER: 08/686,878
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-040-916-50
Query Match 40.0%; Score 200; D3 12; Length 279;
Best Local Similarity 88.2%; Pred. No. 4.6e-60;
Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;
QY 133 ACCAGTCCACCTTTAAACACGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAG 192
DB 1 RCACATCCACCTTTAAACACGGGCTTGCAAAAGGTTACACTTGACCAATCAGAGAG 60
QY 193 CTCACTAAATGCTAATTAGGCAAGAGAGGTTAAAGAAATAGCCATCTATTTC 252
DB 61 NTCANTAAATGATNATTTGGCAAAAACAGAGGTTAAAGAAATAGCCATCTATTTC 120
QY 253 CTGAAGACACACAGAGAGGAGCAACAATGGGATATAACCCAGGATTCGGGTGGGCA 312
DB 121 CTGAAGACACACAGAGAGGAGCAACAATGGGATATAACCCAGGATTCGGGTGGGCA 180
QY 313 CAGCAGCCCCCTTTGGGCTCCCTTTGATGGGAGC--TGTTTCATGCTATTTC 370
DB 181 CGGCA-ACCCCTTTGGGCTCCCTTTGATGGGAGCCTTGTTTCACTGCTATTTC 239
QY 371 CTCTATTAATCTTGCACTGCA 393
DB 240 TMTATTAATTTGCACTGCA 262

RESULT 15
US-09-864-761-20462/c
Sequence 20462, Application US/09864761

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,587
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20462
LENGTH: 246
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010951.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: AB026898.1, EVALUATE 4.00e-57
OTHER INFORMATION: EST_HUMAN HIT: A1492055.1, EVALUATE 3.00e-49
OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALUATE 7.30e-01
US-09-864-761-20462
Query Match 25.2%; Score 126; DB 10; Length 246;
Best Local Similarity 85.6%; Pred. No. 4.5e-34;
Matches 178; Conservative 0; Mismatches 20; Indels 10; Gaps 3;

QY 246 CTATTGCTGAGACACAGAGGGAACAATCGGATATTAACCCAGCATTCGAG 305
|||
Db 246 CTGTTGCTGAGACACAGCGGAGGGAATATATCAAGATATTAACCAAGCATTCGAG 187
|||
QY 306 CTGGCAACAGAGCCCCCTTTGGGTCCCTTTGTTATGGAGCTGTTTCATGCTA 365
|||
Db 186 CTGGCAACGGTA-ACCCCCTTTGGGTCCCTTTGTTATGGAGCTCTAT----- 136
|||
QY 366 TTCTACTTATTAATCTTGCACTGCACTCTCTGGTGCATGTTCTTAAGGCTCGAGC 425
|||
Db 135 CTTTACTCTATTAATCTTGCACTGTACTCTTGGTCCGTTTGTAC-GCTTAGAC 77
|||
QY 426 TGAAGTTTGGCTCACCGTCCACCACTGC 453
|||
Db 76 TGAAGTTTGGCTCGCCATCCACCACTGC 49
|||

Search completed: April 19, 2003, 14:55:18
Job time : 251.566 secs

Accession	AI128526	GI:3597040
KEYWORDS	EST.	
ORGANISM	human.	
REFERENCE	1 (bases 1 to 436)	
AUTHORS	NCI-CCAP	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@b-remail.nih.gov This clone is available royalty-free through LIND ; contact the IMAGE Consortium (info@image.lind.gov) for further information. Insert Length: 720 Seq Error: 0.00 Seq primer: -40m13 fwd. ET from Aterham High quality sequence stop: 428.	
FEATURES	Location/Qualifiers	
SOURCE	1. 436 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1714147" /clone_lib="Soares_placenta:80c9weeks_2NHbPct9w" /dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception" /lab_host="DH10B (ampicillin resistant)" /note="Organ: placenta; Vector: pUT73D (pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTGACCATCTGAGTGAGGAGCGGCGCGCATTTTCTTTTCTTTT 3'], RI double-stranded cDNA was also selected, ligated to Eco RI adapters (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pUT73 vector (pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	95 a 96 c 106 g 139 t	
ORIGIN		
Query Match	56.6%; Score 283.2; DB 9; Length 436;	
Best Local Similarity	95.2%; Pred. No. 7e-12;	
Matches 314; Conservative	0; Mismatches 13; Indels 3; Gaps 2;	
66	ACTGAGAGACAGAGACTAGCTGTGATTTCTCAGGCGGAC	125
351	ACTGGAAGACAGAGACTAGCTGTGATTTCTCAGCTGAC	292
126	GAAAGTGACCACTGCTTTTAAACGCGGGCTTGCACTT	185
291	GAAAGTGACCACTGCTTTTAAACGCGGGCTTGCACTT	232
166	CAGAGAGCTCAGTAAATGCTAATTAGCAAGACGAGGT	245
231	CAGAGAGCTCAGTAAATGCTAATTAGCAAGACGAGGT	172
246	CTATTGCTCAGAGACACAGAGAGGAGCAACATGCG	305
171	CTATTGCTCAGAGACACAGAGAGGAGCAACATGCG	112
306	CTGGCAACAGACACCCCTTTGGGTCCCTTCCCTT	363
111	CCGGCAACAGACACCCCTTTGGGTCCCTTCCCTT	53
364	TATTTCACCTTAAATCTTGCAACTGCA 393	
52	TATTTCACCTTAAATCTTGCAACTGCA 23	
RESULT 6		
LOCUS	AA781423 494 bp mRNA linear EST 31-DEC-1998	
DEFINITION	aj36c03.s1 Soares testis NHT Homo sapiens cDNA clone J391428 3'	

	sequence.	similar to contains PRR7.tl PRR7 repetitive element ; , mRNA
ACCESSION	AA781423	
VERSION	AA781423.1	GI:2840754
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 494)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
JOURNAL	Tumor Gene Index	
COMMENT	Unpublished (1997)	
	Contact: Robert Straubeberg, Ph.D.	
	Email: cgaaps-remail.nih.gov	
	cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo	
	, Ph.D.	
	'Ph.D.	
	DNA Sequencing by: Greg Lennon, Ph.D.	
	Clone distribution: NCI-CGAP clone distribution information can be	
	found through the I.M.A.G.E. Consortium/LNL at:	
	www.bio.lnl.gov/bdrp/image/image.html	
	Insert length: 1645 Std Error: 0.00	
	Seg primer: -40ml3 fwd. EF from Amersham	
	High quality sequence stop: 475.	
FEATURES	Location/Oualifiers	
source	1..494	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/clone="1391428"	
	/clone_1lb="Soares_testis_NHT"	
	/sex="male"	
	/lab_host="DH10B"	
	/note="Vector: pUT7JD-Poc (Pharmacia) with a modified	
	polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA	
	was prepared from mRNA obtained from Clontech Laboratories	
	', Inc.' and primed with a Not I - oligo(dT) primer [5']	
	IGTGAACAATCTGAAGTGAGGAGCGCCGCCAATTITTTTTTTT 3 } .	
	Double-stranded cDNA was ligated to Eco RI adaptors	
	(Pharmacia) , digested with Not I and cloned into the Not I	
	and Eco RI sites of the modified pUT7J vector. Library	
	went through one round of normalization to Cos5 and was	
	constructed by Benito Soares and M. Fatima Bonaldo."	
BASE COUNT	108 a 111 c 133 g 141 t	1 others
ORIGIN		
Query Match	56.4%; Score 282.2; DB 9; Length 494;	
Best Local Similarity	94.8%; Pred. No. 1.6e-81;	
Matches 313; Conservative	0; Mismatches 14; Indels 3; Gaps 2;	
OY	66 ACTGAGAGACAGACTAGCTGATTTTCCTAGCCGACTTAAGAATCCTTAGCCTAGCTGG	125
Db	334 ACTGAGACACAGAGACTAGCTGATTTTCCTAGCTGACTTAAGAATCCTTAGCCTAGCTGG	275
OY	126 GAAGGTGACCAAGTCCACTTTTAAAACAGCGGCTTGCACCTTAGCTCACACTTGACCAAT	185
Db	274 GAAGGTGACCAATCCACTTTTAAAACAGCGGCTTGCACCTTAGCTCACACTTGACCAAT	215
OY	186 CAGAGAGCTCAATAAGCTTAATTAGCCANAAGCAGAGGTAAAGAAATAGCCAATCAT	245
Db	214 CAGAGAGCTCACTTAAATAGCTTAAGCAAAGCAGAGGTAAAGAAATAGCCAATCAT	155
OY	246 CTATTGCTCTGAGACACAGCAGAGGGGCAACAATCGGATATATAACCAGACATTGCG	305
Db	154 CTATTGCTCTGAGACACAGCAGAGGGGCAACAATCGGATATATAACCAGACATTGCG	95
OY	306 CTGGCAACAGACCCCCCTTGGGTCCCTTCCCTTTGTATGGAG-CTGTTTCATGC	363
Db	94 CGGGCAACAGGCA-ACNCCCTTGGGTCCCTTCCCTTTGTGGAGCTGTGTTTCATGC	36
OY	364 TATTTCACCTTATTTAATCTTGCAACTGCA	393

Db 35 TATTCACTCTATTAAATCTTGCAACTGCA 6

RESULT 7
LOCUS AA860368/c 342 bp mRNA linear EST 31-DEC-1998
DEFINITION aj59c05.g1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394600
3' similar to contains PRR7.c1 PTRS repetitive element ;, mRNA
sequence.
AA860368
ACCESSION AA860368.1 GI:2954363
VERSION
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 342)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
COMMENT
CONTACT: Robert Strausberg, Ph.D.
Email: rgsbds-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNL at:
www-bio.lnl.gov/bdip/image/image.html
Insert length: 1305 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 319.
Location/Qualifiers
1..342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394600"
/clone_1bp="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACCATCTGAAGTGGAGGCGGCCCAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library
went through one round of normalization to Cos5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 75 a 78 c 82 g 107 t

ORIGIN

Query Match 56.0%; Score 280; DB 9; Length 342;
Best Local Similarity 94.5%; Pred. No. 7.1e-81;
Matches 312; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

Db 66 ACTGAGAGACGAGACTAGCTGATTTCTAGCGCGACTAAGATTCCTAGCTAGCTGG 125
338 ACTGAGAGACGAGACTAGCTGATTTCTAGCGCTGACTGAATCCCTAAGCTTAGCTGG 279

Db 126 GAAGGTACACGCTCCACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
278 GAAGGTACACATCCACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 219

Db 186 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCAATCAT 245
218 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCAATCAT 159

Db 246 CTATTGCTGAGAGACAGAGAGGAGCAACATCGGAGATTAATCCAGCATTTGAG 305
158 CTATTGCTGAGAGACAGAGAGGAGCAATGATCGGAGATTAATCCAGCATTTGAG 99

Db 306 CTGGACACAGACAGCCCCCTTTGGGTCCTTCCCTTGATGAGAG--CTGTTTCATGC 363
98 GCGGCACAGGCA-AACCCCTTTGGGTCCCTCCCTTTGTATGAGAGCTCGTTTTCATGC 40

Db 364 TATTCACTCTATTAAATCTTGCAACTGCA 393
39 TATTCACTCTATTAAATCTTGCAACTGCA 10

RESULT 8
LOCUS BH149565/c 921 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTQ648TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
BH149565
ACCESSION BH149565.1 GI:15310303
VERSION
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
AUTHORS 1 (bases 1 to 921)
TITLE Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
COMMENT
CONTACT: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 40
High quality sequence stop: 567.
Location/Qualifiers
1..921
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/strain="HMI:IMSS"
/clone_1bp="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1, Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 Kb). The v + 1 method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 242 a 216 c 226 g 237 t

ORIGIN

Query Match 55.4%; Score 277.2; DB 17; Length 921;
Best Local Similarity 84.5%; Pred. No. 9.1e-80;
Matches 377; Conservative 0; Mismatches 48; Indels 21; Gaps 5;

Db 66 ACTGAGAGACGAGACTAGCTGATTTCTAGCGCGACTAAGATTCCTAGCTAGCTGG 125
651 AGTGAAGAGACGAGACTAGCTGATTTCTAGCGCTAAGAAATCTTAAGCTTAGCTGG 592

Db 126 GAAGGTACACGCTCCACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
591 GAAGGTACATCCATCCATTAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 532

Db	Accession	Version	Source	Organism	Definition	LOCUS	Result 9
Oy	186	C	-----	-AAGAGGCTCAGTAAATGCTATTTCGCAAAACAGGAGCTAAAGAAATA	236		
Db	531	CAGTAACTGTAAGAGGCTCAGTAAACGCTTAATTCGCAAAACAGGAGCTAAAGAAATA	472				
Oy	237	GCCAACTCATCTATTGCTCAGAGCAGCAGGAGGAGCAACAATCGGATATTAACCCAG	296				
Db	471	GCCAACTCATCTATTGCTCAGAGCAGCAGGAGGAGCAACAATCGGATATTAACCCAG	413				
Oy	297	GCATTCCGAGCTGGCAACAGACCCCCCTTGGGTCCTTCCCTTGGATATGGAGACTGTT	356				
Db	412	GCATTCCGAGCTGGCAACAGACCCCCCTTGGGTCCTTCCCTTGGATATGGAGACTGTT	354				
Oy	357	TTCATGCTATTTCACCTCTATTAAATCTTGCACACTGCTCTTCGCTCATGTTTCTT	414				
Db	353	T-----TTTCACCTCTATTAAATCTTGCACACTGCTCTTCGCTCATGTTTCTT	302				
Oy	415	ACGGCTCCAGCTGAGCTTTTGCTCACCGCTCCACACACTGCTGTTTGGCACACCCGAGACC	474				
Db	301	ACGACTCCAGCTGAGCTTTTGCTCACCGCTCCACACACTGCTGTTTGGCACACCCGAGACC	242				
Oy	475	TGCGCGTGAAGCTCCCAATCCCTCTGAT	500				
Db	241	CGCGCGTGAAGCTCCCAATCCCTCTGAT	216				
Result 9	N53177/c						
LOCUS	N53177						
DEFINITION	Y566h1.81 Soares fetal liver spleen INFLS Homo sapiens cDNA clone	422 bp	mRNA	linear	EST 28-JAN-1997		
KEYWORDS	IMAGE:246789.3						
SOURCE	EST.						
ORGANISM	human.						
REFERENCE	Hukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
AUTHORS	Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.						
TITLE	1 (bases 1 to 422)						
JOURNAL	Hillier, L., Lennon, G., Becker, M., Bernaldo, M.F., Chiapelli, B.,						
MEDLINE	Chisoso, S., Dietrich, N., Dubuque, T., Favellio, A., Gish, M., Hawkins						
COMMENT	'M., Hillman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore						
	'B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,						
	Schellander, K., Wohlmann, P., Wetzelsch, R., Wilson, R. and Warr, M.						
	Underwood, K., Wohlmann, P., Wetzelsch, R., Wilson, R. and Warr, M.						
	Generation and analysis of 280,000 human expressed sequence tags						
	Genome Res. 6 (9), 807-828 (1996)						
	97044478						
	Contact: Wilson RK						
	Washington University School of Medicine						
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108						
	Tel: 314 286 1800						
	Fax: 314 286 1810						
	Email: est@watson.wustl.edu						
	This clone is available royalty-free through INFL; contact the						
	IMAGE Consortium (info@image.lnl.gov) for further information.						
	Insert Length: 938						
	Seq Error: 0.00						
	Seq primer: m3 -40 forward						
	High quality sequence stop 301.						

```

FEATURES
source
location/Qualifiers
1..422
/organism="Homo sapiens"
/db_xref="GDB:3796035"
/db_xref="taxon:9606"
/clone="IMAGE:246789"
/clone_1ib="Scares fetal liver spleen INFRs"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="PH10B (ampicillin resistant)"
note="Organ: Liver and Spleen; Vector: pT7TD (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer (5' AACGCGAAGATTAATTAACATCTTTTTTTTTTTTTTTTTTTT 3'),

```

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) , digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."	BASE COUNT
95 a	95 c
111 g	120 t
1 others	
ORIGIN	

Query Match	Similarity	Score	DB	Length
Best Local Similarity	93.3%	Pred. No. 1, 66-78;		
Matches	308;	Conservative	0;	Mismatches 16; Indels 6; Gaps 2;

QY	DB	Sequence	Score
QY	66	ACTGAGACAGACGACTAGCTGATTTCTCTAGCCGACATAAGATCTTAAGCTTAGCTGG	125
DB	327	ACTGAGACAGACGACTAGCTGATTTCTCTAGCCGACATAAGATCTTAAGCTTAGCTGG	268
QY	126	GAAGGTGACACAGTCCACCTTTAAACACAGGGGCTTGCACTTAGTCAACCTGACCAAT	185
DB	267	GAAGGTGACACAGTCCACCTTTAAACAGGGGCTTGCACTTAGTCAACCTGACCAAT	208
QY	186	CAGAGAGCTCATTAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT	245
DB	207	CAGAGAGCTCATTAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT	148
QY	246	CTATTGCTTGAGACACAGCAGAGGGGCAACAAATCGGGATATATAAACCAGCATTTGAG	305
DB	147	CTATTGCTTGAGACACAGCAGAGGGGCAACAAATGATCGGGATATATAAACCAGCATTTGAG	88
QY	306	CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCTTTGATGGGAG--CTGTTTTCATGC	363
DB	87	CGG----CAAGGGCAACCCCTTTGGTCCCTTCCTTTGATGGGAGCTGTGTTTCATGC	32
QY	364	TATTTCACTCTATTAAATCTTGCAACTGCA	393
DB	31	TATTTCACTCTATTAAATCTTGCAACTGCA	2

RESULT	10
H01325/c	
LOCUS	
DEFINITION	H01325 388 bp mRNA linear EST 19-JUN-1995
IMAGE	y19ge01.s1 Soares placenta N22HP Homo sapiens cDNA clone
VERSION	167384_3', mRNA sequence.
KEYWORDS	H01325
SOURCE	H01325.1 GI:864258
ORGANISM	EST.
REFERENCE	human.
AUTHORS	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo. 1 (bases 1 to 388) Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman, 'M.', Hultman,M., Kucaba,T., Le'M., Lennon,G., Maira,M., Parsons,J., Rifkin,T., Rohlfing,T., Soares,M., Tan,F., Trevaethis,E., Waterston , R., Williamson,A., Woldmann,P. and Wilson,R. The Mashu-Merck EST Project Unpublished (1995) Contact: Wilson RK
TITLE	
JOURNAL	
COMMENT	

Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 790
High quality sequence strops: 346
Source: IMAGE Consortium, LNLN
This clone is available royalty-free through LNLN ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 790 Std Error: 0.00
Seq primer: Promega -2im13
High quality sequence strop: 346.
location/Qualifiers
1..388
source
organism="Homo sapiens"

/db_xref="GDB:559031"
 /db_xref="taxon:9606"
 /clone="IMAGE:147384"
 /clone_1ib="Soares placenta Nb2HP"
 /sex="Female"
 /dev_stage="Placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', ACTGGAAGAATTGCGCGCGAGAAATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 90 a 90 c 96 g 110 t 2 others

ORIGIN

Query Match 54.4%; Score 271.8; DB 14; Length 388;
 Best Local Similarity 93.8%; Pred. No. 3.8e-78;
 Matches 305; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAGATCCCTAGCTGCTG 125
 DB 329 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAGATCCCTAGCTGCTG 270
 QY 126 GAGGTGACCAACGCTCCCTTTAAACAGGGGCTTGCACTTACCTACACCTGACCAAT 185
 DB 269 GAGGTGACCAACGCTCCCTTTAAACAGGGGCTTGCACTTACCTACACCTGACCAAT 210
 QY 186 CAGAGAGCTCACTAAATGCTATTAGGCAAGAAGAGAGGTAAGAATAGCAATCAT 245
 DB 209 CAGAGAGCTCACTAAATGCTATTAGGCAAGAAGAGAGGTAAGAATAGCAATCAT 150
 QY 246 CTATTGCTTGAAGACAGACAGAGGCAACAATCGGATATTAACCCAGGCAATTCGAG 305
 DB 149 CTATTGCTTGAAGACAGACAGAGGCAACAATCGGATATTAACCCAGGCAATTCGAG 90
 QY 306 CTGGCAACAGAGCGCCCTTTGGGCTCCCTTTGTATGGAG--CTGTTTTCATGC 363
 DB 89 GCGGCAACGGCA-AGCCCCCTTTGGGCTCCCTTTGTATGGAGCTGTGTTTTCATGC 31
 QY 364 TATTCACTCTATTAATCTTGCA 388
 DB 30 TATTCACTCTATTAATCTTGCA 6

RESULT 11 385 bp mRNA linear EST 28-JAN-1997
 N55091/C
 LOCUS YV43803.81 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 DEFINITION IMAGE:245500 3, mRNA sequence.
 ACCESSION N55091
 VERSION N55091.1 GI:1197970
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gibb, M., Hawkins, M., Hultman, M., Kucada, T., Lacy, M., Le, M., Le, N., Marchis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

TITLE JOURNAL
 MEDLINE
 COMMENT
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through INFL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 907 Std Error: 0.00
 Seq primer: m13 -40 forward
 High quality sequence stop: 276.
 Location/Qualifiers
 1. 385

/organism="Homo sapiens"
 /db_xref="GDB:3794746"
 /db_xref="taxon:9606"
 /clone="IMAGE:245500"
 /clone_1ib="Soares fetal liver spleen INFLS"
 /sex="Male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: liver and spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', ACTGGAAGAATTATTAAGAATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 88 a 88 c 102 g 106 t 1 others

ORIGIN

Query Match 54.2%; Score 271.2; DB 14; Length 385;
 Best Local Similarity 93.0%; Pred. No. 5.9e-78;
 Matches 306; Conservative 0; Mismatches 19; Indels 4; Gaps 2;

QY 66 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAGATCCCTAGCTGCTG 125
 DB 327 ACTGAGAGACGAGCTAGCTGATTTCTTACGCGACTAGATCCCTAGCTGCTG 268
 QY 126 GAGGTGACCAACGCTCCCTTTAAACAGGGGCTTGCACTTACCTACACCTGACCAAT 185
 DB 267 GAGGTGACCAACGCTCCCTTTAAACAGGGGCTTGCACTTACCTACACCTGACCAAT 208
 QY 186 CAGAGAGCTCACTAAATGCTATTAGGCAAGAAGAGAGGTAAGAATAGCAATCAT 245
 DB 207 CAGAGAGCTCACTAAATGCTATTAGGCAAGAAGAGAGGTAAGAATAGCAATCAT 148
 QY 246 CTATTGCTTGAAGACAGACAGAGGCAACAATCGGATATTAACCCAGGCAATTCGAG 305
 DB 147 CTATTGCTTGAAGACAGACAGAGGCAACAATCGGATATTAACCCAGGCAATTCGAG 88
 QY 306 CTGGCAACAGAGCGCCCTTTGGGCTCCCTTTGTATGGAG--CTGTTTTCATGC 363
 DB 87 C-GGCAACGGCAACCCCTTTGGGCTCCCTTTGTATGGAGCTGTGTTTTCATGC 30
 QY 364 TATTCACTCTATTAATCTTGCACTGC 392
 DB 29 TATTCACTCTATTAATCTTGCACTGC 1

RESULT 12. 342 bp mRNA linear EST 01-FEB-1995
 T47345/C
 LOCUS YD10802.81 Stragagene placenta (#937225) Homo sapiens cDNA clone
 DEFINITION IMAGE:70803 3, mRNA sequence.
 ACCESSION T47345
 VERSION T47345.1 GI:649326
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,

Chisoe, S., Dietrich, N., Dubuque, T., Favella, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry, M. J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Matra, M. Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

TITLE
JOURNAL
MEDLINE

COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 598

Source: IMAGE Consortium, LNL. This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 598 Std Error: 0.01
Seq primer: -21ml3

High quality sequence stop: 343.

FEATURES
Location/Qualifiers

Source

1. .342

/organism="Homo sapiens"

/db_xref="GDB:491700"

/db_xref="taxon:9606"

/clone="IMAGE:70803"

/clone_lib="Stratagene placenta (#937225)"

/sex="male"

/lab_host="SOLR cells (kanamycin resistant)"

/note="Organ: placenta; Vector: pBluescript SK-; Site: 1; EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Caucasian. Average insert size: 1.2 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GATTGGGACGAG 3' -3' adaptor sequence: 5' CTCGACTTTTCTTTTCTTTT 3'."

BASE COUNT 74 a 80 c 80 g 102 t 6 others

ORIGIN

Query Match 54.2%; Score 270.8; DB 14; Length 342;
Best Local Similarity 92.1%; Pred. No. 7.6e-78;
Matches 304; Conservative 0; Mismatches 23; Indels 3; Gaps 2;

QY 66 ACTGAGACAGAGCTAGCTGATTTCTTACGCGCACTAAGATCCCTAAGCTAGCTGG 125
DB 331 ACTGAGACAGAGCTAGCTGATTTCTTACGCGCACTAAGATCCCTAAGCTAGCTGG 272
QY 126 GAAGTGACACGCTCCTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAAT 185
DB 271 GAAGTGACACGCTCCTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAAT 212
QY 186 CAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGTAAGTAAGTACCAATCA 245
DB 211 CAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGTAAGTAAGTACCAATCA 152
QY 246 CTATTGCTGAGAGCAGAGAGAGAGCAACATCGATATTAACCCAGGCAATTCGAG 305
DB 151 CTATTGCTGAGAGCAGAGAGAGAGCAACATCGATATTAACCCAGGCAATTCGAG 92
QY 306 CTGGCAACAGAGCCCCCTTTGGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTT 363
DB 91 GCGGCAACAGGCA-ANCCCCCTTTGGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCT 33
QY 364 TATTTCACCTTAAATTTGCAACTGCA 393
DB 32 TATTTCACCTTAAATTTGCAACTGCA 3

RESULT 13
LOCUS B1087886 771 bp mRNA linear EST 20-JUN-2001
DEFINITION 602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4993894 5',
ACCESSION B1087886
mRNA sequence.

VERSION B1087886.1 GI:14506216
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 771)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-f@mail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: LLM11015 row: d column: 23

High quality sequence stop: 762.

FEATURES
Location/Qualifiers

Source

1. .771

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4993894"

/clone_lib="NIH_MGC_10"

/cell_line="MGC36"

/lab_host="DH108"

/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.5 kb. Library prepared by Life Technologies."

BASE COUNT 225 a 206 c 166 g 174 t

ORIGIN

Query Match 54.1%; Score 270.6; DB 13; Length 771;
Best Local Similarity 94.6%; Pred. No. 1.3e-77;
Matches 313; Conservative 0; Mismatches 14; Indels 4; Gaps 3;

QY 66 ACTGAGACAGAGCTAGCTGATTTCTTACGCGCACTAAGATCCCTAAGCTAGCTGG 125
DB 429 ACTGAGACAGAGCTAGCTGATTTCTTACGCGCACTAAGATCCCTAAGCTAGCTGG 488
QY 126 GAAGTGACACGCTCCTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAAT 185
DB 489 GAAGTGACACGCTCCTTAAACAGGGGCTTGCACTTAGCTCACCTGACCAAT 548
QY 186 CAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGTAAGTAAGTACCAATCA 244
DB 549 CAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGTAAGTAAGTACCAATCA 608
QY 245 TCTATTGCTGAGAGCAGAGAGAGAGCAACATCGGATATTAACCCAGGCAATTCGAG 304
DB 609 TCTATTGCTGAGAGCAGAGAGAGAGCAACATCGGATATTAACCCAGGCAATTCGAG 668
QY 305 GCTGGCAACAGAGCCCCCTTTGGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCTT 362
DB 669 GCGGCAACAGGCA-ACCCCCCTTTGGGCTCCCTTCTTCTTCTTCTTCTTCTTCTTCT 727
QY 363 CTATTTCACCTTAAATTTGCAACTGCA 393
DB 728 CTATTTCACCTTAAATTTGCAACTGCA 758

RESULT 14
LOCUS AW971553 522 bp mRNA linear EST 01-JUN-2000
DEFINITION EST383642 MAGE resequences, MAGL Homo sapiens cDNA, mRNA sequence.
ACCESSION AW971553
VERSION AW971553.1 GI:8161399
KEYWORDS EST.
SOURCE human.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Human sapiens					
Euryarchaeota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
1 (bases 1 to 522)					
Hedge, P., Qi, R., Abernathy, K., Dharp, S., Gaspar, R., Gay, C., Holt, I. E., Saeed, A. I., Sharov, V., Lee, N. H., Yeatman, T. J. and Quackenbush, J.					
Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray					
unpublished (2000)					
Contact: John Quackenbush					
The Institute for Genomic Research					
9712 Medical Center Dr., Rockville, MD 20850, USA					
Tel: 301 838 3528					
Fax: 301 838 0208					
Email: john@tigr.org					
Plate: 292					
Seq primer: Forward.					
Location/Qualifiers					
1..522					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/clone_id="MAGE_researches_MAGI"					
/note="Vector: pBluescriptSM"					
BASE COUNT	125 a	115 c	134 g	148 t	
ORIGIN					
Query Match	53.5%	Score 267.4;	DB 10;	Length 522;	
Best Local Similarity	92.7%	Pred. No. 1.2e-76;			
Matches 303; Conservative	0;	Mismatches 21;	Indels 3;	Gaps 2;	
QY	66	ACTGAGACAGCAGCAGCTAGCTGATTTCTTAGGCCGACCTAAGATCCCTTAAGCTTACCTG	125		
Db	326	ACTGAGACAGCAGCAGCTAGCTGATTTCTTAGGCCGACCTAAGATCCCTTAAGCTTACCTG	267		
QY	126	GAAAGTGACCAAGTCACCTTTAAACAGGGGGCTTGCAACTAGCTCAGCAGCTGACCAAT	185		
Db	266	GAAAGTGACCGATCACCCTTTAAACAGGGGGCTTGCAACTAGCTCAGCAGCTGACCAAT	207		
QY	186	CAGAGAGCTCACTAAATAGCTAATTAGGCMAAGACAGAGCTTAAGAAATAGCAATCAT	245		
Db	206	CAGAGAGCTCACTAAATAGCTAATTAGGCMAAGAAACAGAGCTTAAGAAATAGCAATCAT	147		
QY	246	CTATTGCTGAGAGCAGCAGAGGGAGCACAATCGGAGATTAATTAACCCAGGCAATTCGAG	305		
Db	146	CTATTGCTGAGAGCAGCAGAGGGAGCACAATCGGAGATTAATTAACCCAGGCAATTCGAG	87		
QY	306	CTGGCAACAGCAGCCCCCTTTGGGTCCCTTCCTTGATAGGAG--CTGTTTCATGC	363		
Db	86	CCAGCAACGGCA-ACCGCTTTGGGTCCCTTCCTTGATAGGAGCTGTTTCATGC	28		
QY	364	TATTTCACTCTAATTAATCTTCAACT	390		
Db	27	TATTTCACTCTAATTAATCTTCAACT	1		
RESULT 15					
LOCUS	T69704	363 bp	mRNA	linear	EST 07-MAR-1995
DEFINITION	Y013A03.81 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:66988 3', mRNA sequence.				
ACCESSION	T69704				
VERSION	T69704.1	GI:680852			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 363)				
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaaskis, B., Waterston, R., Williamson, A., Wohldmann, F., and Wilson, R.				

```

TITLE      The MashU-Merck EST Project
JOURNAL    Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel.: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu

Insert Size: 766
High quality sequence stops: 341 Source: IMAGE Consortium, LINTL This clone is available royalty-free through LINTL ; contact the IMAGE Consortium (info@image.lnll.gov) for further information.
Insert length: 766 Std Error: 0.00
Seq primer: -21m13
High quality sequence stop: 341.
Location/Qualifiers
1 . 363
/drganism="Homo sapiens"
/db_xref="GDB:463693"
/db_xref="tixon:9606"
/c1one="IMAGE:66988"
/clon_e_1db="Soares fetal liver spleen INFLS"
/sex="male"
/dev stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTR73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGCAGCATTAATTTAAAGAATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTR73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT   83 a          83 c          87 g        107 t         3 others
ORIGIN
Query Match           53.0%; Score 265; DB 14; Length 363; Best Local Similarity 89.6%; Pred. No. 6.3e-76; Matches 294; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY  ACTGAGACAGCAGCACTAAGATTTCCTTAGCCGACAATAATGCCGAAGCTAGCTGG 125
Db  ACTGAGACAGCAGCACTAAGATTTCCTTAGCGACGTAAATATCCCAGACCTAGCTGG 125
QY  GAAGGTGACCACGTCACCTTTAAAACAGGGGGCTTGCAAATTAAGCTCACACCTGACCAAT 185
Db  GAAGGTGACCACATCCACTTTAAAACAGGGGGCTTGCAAATTAAGCTCACACCTGACCAAT 209
QY  CAGAGAGCTCCTAAATGCTAAATTTAGCCAAGACAGAGGTAAAGAATATGCCATATAT 245
Db  CAGAGAGCTCCTAAATGCTAAATTTAGCCAAGACAGAGGTAAAGAATATGCCATATAT 149
QY  CTATTGCGCTGAGACAGCAGGAGGAGCAACATCGGGATATATAACCCAGACATTTGAG 305
Db  CTATTGCGCTGAGACAGCAGGAGGAGCAATGATCGGGATATATAACCCAGATCTTCGAN 89
QY  CTCGACACAGCAGCCCCCCTTTGGSGTCCTTCCTTTGATATGGAGAGCTGTTTTCATGCTA 365
Db  CGGACAGGCAACCCCTTTGGGGCCCTCCCTTTGATATGGAGAGCTGTTTTCATGCTA 30
QY  TTTCACCTCTATTAAATCTTGCAACTGCA 393
Db  TTTCACCTCTATTAAATCTTNNAAGNACA 2

```

Search completed: April 19, 2003, 14:46:28
Job time : 1271.98 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:28:30 ; Search time 1700.07 Seconds
(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500
Perfect score: 500
Sequence: 1 cctctggggggggcctctctt.....tgactccatccctctcgat 500.

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pl: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	10499	6	AX007980	AX007980 Sequence
2	500	100.0	56093	6	AX329572	AX329572 Sequence
3	500	100.0	56093	6	HSAC000064	AC000064 Human BAC
4	500	100.0	149194	9	AC007566	AC007566 Homo sapi
5	435	87.0	711	6	AX007997	AX007997 Sequence
6	410	82.0	783	6	AX000970	AX000970 Sequence
7	383	76.8	711	6	AX007998	AX007998 Sequence
8	360	72.0	137482	2	AL161721	AL161721 Homo sapi
9	359.8	72.0	143590	9	HS8363	AL035706 Human DNA
10	358.2	71.6	119406	2	AC103595	AC103595 Homo sapi
11	357.8	71.6	110000	2	AL353584_0	AL353584 Homo sapi
12	357.8	71.6	169029	9	CNS06C7R	AL390800 Human chr
13	357.8	71.6	170746	9	AL392173	AL392173 Human DNA
14	357.4	71.5	163803	9	AC093531	AC093531 Homo sapi
15	357.4	71.5	191863	9	AC010888	AC010888 Homo sapi
16	356.8	71.4	180635	9	AL360169	AL360169 Human DNA
17	356.6	71.3	174019	9	AP001538	AP001538 Homo sapi
18	356.6	71.3	340000	9	AP001674	AP001674 Homo sapi
19	354.2	70.8	111079	9	HS419C19	AL036407 Human DNA
20	353.4	70.7	76169	9	AC003014	AC003014 Human PAC
21	351.8	70.4	161049	2	AL591842	AL591842 Homo sapi
22	351	70.2	42216	9	AC092899	AC092899 Homo sapi
23	351	70.2	170414	2	AC026019	AC026019 Homo sapi
24	349.6	69.9	83412	9	AC092843	AC092843 Homo sapi
25	349.2	69.8	165059	9	AC106856	AC106856 Homo sapi
26	348.2	69.6	839	9	AF127226	AF127226 Homo sapi
27	348.2	69.6	1329	6	AX001030	AX001030 Sequence
28	347.6	69.5	177212	2	AC025733	AC025733 Homo sapi
29	347.6	69.5	181753	9	AC124075	AC124075 Homo sapi
30	347.2	69.4	119481	9	AL136234	AL136234 Human DNA
31	347.2	69.4	152986	9	AC013719	AC013719 Homo sapi
32	346.4	69.3	179585	2	AC024977	AC024977 Homo sapi
33	346.4	69.3	185671	9	AC084193	AC084193 Homo sapi
34	346.2	69.2	156177	9	AC093896	AC093896 Homo sapi
35	345	69.0	37674	9	HSU221F2	275746 Human DNA
36	345	69.0	143620	9	AC117425	AC117425 Homo sapi
37	344.8	69.0	157711	9	AC055738	AC055738 Homo sapi
38	344.8	69.0	181850	9	AC018680	AC018680 Homo sapi
39	344.8	69.0	207945	9	CNS01D5S	AL121841 Human chr
40	344.8	69.0	318442	2	AC021652	AC021652 Homo sapi
41	344.6	68.9	140756	9	AL139038	AL139038 Human DNA
42	344.6	68.9	183499	2	AL607153	AL607153 Homo sapi
43	344.4	68.9	88328	9	AL357874	AL357874 Human DNA
44	343	68.6	187321	2	AC092510	AC092510 Papio cyn
45	341.6	68.3	102200	9	HS30M3	AL031775 Human DNA

ALIGNMENTS

RESULT 1
LOCUS AX007980 10499 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO9667395.
ACCESSION AX007980
VERSION AX007980.1 GI:9995677
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 10499)
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

Pred. No. is the number of results predicted by chance to have a

JOURNAL

Patent: WO 9967395-A 3 29-DEC-1993;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIET, PATRICK M (FR)

FEATURES

Location/Qualifiers
1..10499
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8.5e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 60
Db 1 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 60
Qy 61 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAG;CCGACTAAGAAATCCCTAAGCCTA 120
Db 61 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAG;CCGACTAAGAAATCCCTAAGCCTA 120
Qy 121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGCTTGCACTTACTGCTACACCTGA 180
Db 121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGCTTGCACTTACTGCTACACCTGA 180
Qy 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAT;GACAGAGGTAAAGAAATAGCCA 240
Db 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAT;GACAGAGGTAAAGAAATAGCCA 240
Qy 241 ATCATCTATTGCTGAGAGACAGAGAGAGGACAAATCGGGATATTAACCCAGGCAAT 300
Db 241 ATCATCTATTGCTGAGAGACAGAGAGAGGACAAATCGGGATATTAACCCAGGCAAT 300
Qy 301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy 361 TGCTATTTCATCTATTAAATCTTGCACTGACTCT;CTGGTCAATGTTTCTTAAGGCT 420
Db 361 TGCTATTTCATCTATTAAATCTTGCACTGACTCT;CTGGTCAATGTTTCTTAAGGCT 420
Qy 421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTT;GCCACCAACCGAGAGCTGCCG 480
Db 421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTT;GCCACCAACCGAGAGCTGCCG 480
Qy 481 TGACTCCCATCCCTCTGTGAT 500
Db 481 TGACTCCCATCCCTCTGTGAT 500
```

RESULT 2
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 81 from Patent WO0194629.
DEFINITION AX329572
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G.,
Hortigian, S., Soper, D.R. and Weaver, J.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
Avalon Pharmaceuticals (US)

FEATURES
SOURCE 1..56093
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 56093;
Best Local Similarity 100.0%; Pred. No. 9.4e-155;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 60
Db 28001 CCTGTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAAGCCCTGGAGATACAGCAATTATC 28060
Qy 61 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAG;CCGACTAAGAAATCCCTAAGCCTA 120
Db 28061 TTGCAACTGAGAGACAGAGACTAGCTGATTTCTAG;CCGACTAAGAAATCCCTAAGCCTA 120
Qy 121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGCTTGCACTTACTGCTACACCTGA 180
Db 28121 GCTGGGAAGGTGACCAAGCTCCACCTTTAAACAGGGGCTTGCACTTACTGCTACACCTGA 180
Qy 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAT;GACAGAGGTAAAGAAATAGCCA 240
Db 28181 CCAATCAGAGAGCTCACTAAATGCTAATTAGCAAT;GACAGAGGTAAAGAAATAGCCA 28240
Qy 241 ATCATCTATTGCTGAGAGACAGAGAGAGGACAAATCGGGATATTAACCCAGGCAAT 300
Db 28241 ATCATCTATTGCTGAGAGACAGAGAGAGGACAAATCGGGATATTAACCCAGGCAAT 28300
Qy 301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 28301 TCGAGCTGGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 28360
Qy 361 TGCTATTTCATCTATTAAATCTTGCACTGACTCT;CTGGTCAATGTTTCTTAAGGCT 420
Db 28361 TGCTATTTCATCTATTAAATCTTGCACTGACTCT;CTGGTCAATGTTTCTTAAGGCT 28420
Qy 421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTT;GCCACCAACCGAGAGCTGCCG 480
Db 28421 CGAGCTGAGCTTTTGTCTACCGCTCCACCACTGCTGTT;GCCACCAACCGAGAGCTGCCG 28480
Qy 481 TGACTCCCATCCCTCTGTGAT 500
Db 28481 TGACTCCCATCCCTCTGTGAT 28500
```

RESULT 3
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
LOCUS Human BAC clone RG083M05 From 7q21-7q22, complete sequence.
DEFINITION AC000064
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 56093)
AUTHORS Paulley, A.
TITLE The sequence of H. sapiens BAC clone RG083M05
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 56093)
AUTHORS Wazerston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: sapiens@wustl.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBEO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SMS1725.

Location/Qualifiers

1. 56093

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone_1b="H_RG083M05"

/clone_1b="CITB-978SK-B"

/complement(838..1131)

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(<1360..1503,4181..4370,4587..4774,6422..6556,

9483..9547,11631..11773,11864..12021,13131..13296,

14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="Atpase; strong similarity to peroxisome

biosynthesis protein PAB1 (PID:g1172019); coded for by

human cDNA C04279 (NID:g1467550)"

/codon_start=1

/protein_id="AAB46346.1"

/db_xref="GI:1669371"

/translation="KRLNIQKTELEVAESEAVMQSVLLDDLDLIALGPAVEHEH

SPAEORCEILCNVIRKMKDCKINFTDLDQVAKETGFVARDVLAIRHLR

SRQISIRKELVLTIDFQKALRGRLPASLRSLNLRKPRDGLGGAEHVQIIMD

TIIDPAKVCCKKKEPELPIANDLPIRQRTGLILGPRGTGTLACVTAARSRNPIVS

KGBELSKYTGASEQAVRDIFRAQAKPCILFDEFESIAPRGHDTGVTVVNO

LLTQLDGVEGLQGVYVLAATSRPDLIDPALIRGRDKCYCPDPQVITISLESKTQ

QMHSFLVSRLEILNIVLSDPLADVDLQHVASVDSFGADLKALVYAQLALHG

MLSKNSEILPDSKEMRYRLPGSSYESIELGNGTSDLSGLSAPSSWTODLPGVP

GKQDPSQPRVLRRTASQEGCELQEGRDRLRADIISIKRIRYSGSGEDSMNQPPRI

KTRLAISQSHLMTRALGHTIRPSISDDKNAEL"

/complement(4948..5130)

/rpt_family="ALU"

/complement(6581..7133)

/rpt_family="U1"

/complement(7767..8037)

/rpt_family="ALU"

/complement(8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:g1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:g1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N22627 (NID:g1130501), bases

276-343"

repeat_region

/rpt_family="ALU"

complement(12612..12907)

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

143-266"

13794..13877

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

30-58"

13907..14104

/rpt_family="ALU"

complement(14110..14137)

/rpt_family="U1"

complement(15618..15907)

/rpt_family="ALU"

17227..17522

/rpt_family="ALU"

16667..19235

/note="match to human fetal brain 5' EST D61494

(NID:g970409), bases 1-255, and to human 3' EST R07476

(NID:g758399)"

19550..19670

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polyproteins;

pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:g842257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

complement(38938..39224)

/rpt_family="ALU"

39225..39707

/note="match to multiple human ESTs, see N30113

(NID:g148633)"

39800..40085

/rpt_family="ALU"

complement(40247..40538)

/rpt_family="ALU"

complement(40632..40924)

/rpt_family="ALU"

complement(42283..42891)

/rpt_family="ALU"

complement(45474..45613)

/rpt_family="ALU"

complement(45614..45737)

/note="match to human 3' EST H48898 (NID:g988738), bases

129-333"

complement(46107..47026)

/note="match to multiple human ESTs, see N81064

(NID:g1243765), H48897 (NID:g988737), and M78831

(NID:g273146)"

complement(47027..47318)

/rpt_family="ALU"

complement(47365..47782)

/note="match to multiple human ESTs, see W37495

(NID:g1319089)"

47898..48115

/note="match to human 5' EST H62306 (NID:g1015138), bases

93-368"

complement(48116..48405)

/rpt_family="ALU"

complement(48406..48584)

/note="match to human 3' EST N29952 (NID:g1148472), bases

290-455, and 5' EST R12730 (NID:g765806)"

complement(48787..49405)

/rpt_family="ALU"

complement(49406..49534)

/note="match to human 3' EST R65794 (NID:g838432), bases

309-440"

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBel0BAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```
1.149194
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_1lb="CTB-978SK-B"
1.1634
/rpt_family="L1"
misc_feature
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
misc_feature
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
repeat_region
2696..3066
/rpt_family="MAIR"
repeat_region
3108..3392
/rpt_family="A1u"
misc_feature
3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
3542..3628
/note="match to EST BG766882 (NID:g14077535)"
```

```
misc_feature
3542..3628
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
3542..3609
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
3562..3628
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
3598..3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature
3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature
3717..3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature
3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
3719..3785
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
3719..3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
3719..3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature
3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
3719..3785
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
3719..3785
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature
3860..3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region
4730..4827
/rpt_family="L2"
repeat_region
4834..4925
/rpt_family="L1"
misc_feature
5715..5860
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
5715..5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
5715..5860
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
5715..5860
/note="match to EST BE299708 (NID:g9183456)"
misc_feature
5715..5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature
5715..5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
5715..5860
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
5715..5860
/note="similar to Mus musculus EST BE994936
```

Query Match 100.0%; Score 500; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 1e-154;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CCTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAACGCTTGAGTAAAGCAATTATC 60
DB 94003 CCTGGGGGGGGCTTCTTCTTGGGATGAGGGCAAAACGCTTGAGTAAAGCAATTATC 93944
QY 61 TTGCACTGAGAGACAGACTAGTGATTTCTTAGGCGAGCTAAGAAATCCCTAAGCTTA 120
DB 93943 TTGCACTGAGAGACAGACTAGTGATTTCTTAGGCGAGCTAAGAAATCCCTAAGCTTA 93884
QY 121 GCTGGAGAGGTGACACAGTCACTTTAAACAGCGGGCTTGCACTTAAAGTCAACCTGA 180
DB 93883 GCTGGAGAGGTGACACAGTCACTTTAAACAGCGGGCTTGCACTTAAAGTCAACCTGA 93824
QY 181 CCAATCAGAGAGTCACTAATAATGCTTAATTAGGCAAGAGAGAGTAAAGAAATAGCCA 240
DB 93823 CCAATCAGAGAGTCACTAATAATGCTTAATTAGGCAAGAGAGAGTAAAGAAATAGCCA 93764
```

Qy 241 ATCATCTATGCTGAGACAGACAGAGGAGGACATCATCGGATATATAACAGAGCAT 300
Db 93763 ATCATCTATGCTGAGACAGACAGAGGAGGACATCATCGGATATATAACAGAGCAT 93704
Qy 301 TCGAGCTGGCAACAGACAGACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCA 360
Db 93703 TCGAGCTGGCAACAGACAGACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCA 93644
Qy 361 TGCATTTTCACTTAATTAATCTTGGCACTGACCTCTGTGTCATGTTCTTAAGGCT 420
Db 93643 TGCATTTTCACTTAATTAATCTTGGCACTGACCTCTGTGTCATGTTCTTAAGGCT 93584
Qy 421 CGAGCTGAGCTTTTGCTACCGCTCCACCACTGCTGTGTGCCACACCGCAGACTGCCGC 480
Db 93583 CGAGCTGAGCTTTTGCTACCGCTCCACCACTGCTGTGTGCCACACCGCAGACTGCCGC 93524
Qy 481 TGAATCCCATCTCTCTGAT 500
Db 93523 TGAATCCCATCTCTCTGAT 93504

RESULT 5
AX007997 711 bp DNA linear PAT 06-SEP-2000
LOCUS Sequence 20 from Patent WO967395.
DEFINITION AX007997
ACCESSION AX007997.1 GI:995694
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 711)
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL Patent: WO 967395-A 20 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
source 1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 170 a 204 c 162 g 175 t
ORIGIN

Query Match 87.0%; Score 435; DE 6; Length 711;
Best Local Similarity 100.0%; Pred. No. 3.2e-133;
Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCCCACTAAGATCCCTAAGCTGCTGG 125
Db 1 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCCCACTAAGATCCCTAAGCTGCTGG 60
Qy 126 GAAGGTGACAGCTGCTCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACAT 185
Db 61 GAAGGTGACAGCTGCTCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACAT 120
Qy 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGTAAAGAAATAGCCATCAT 245
Db 121 CAGAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGTAAAGAAATAGCCATCAT 180
Qy 246 CTAATGCTGAGAGCAGACAGAGAGGAGCAACAATGGCAATTAACCCAGGATTCGAG 305
Db 181 CTAATGCTGAGAGCAGACAGAGAGGAGCAACAATGGCAATTAACCCAGGATTCGAG 240
Qy 306 CTGGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCAATGCTA 365
Db 241 CTGGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCAATGCTA 300
Qy 366 TTTCACTTAATTAATCTTGCAACTGCACTCTTCTGT CATGTTTCTTAAGGCTGAGC 425
Db 301 TTTCACTTAATTAATCTTGCAACTGCACTCTTCTGT CATGTTTCTTAAGGCTGAGC 360

Qy 426 TGAAGCTTTGCTCACCGCTCCACCACTGCTGTTGGCCACACCGCAGACTGCGCTGACT 485
Db 361 TGAAGCTTTGCTCACCGCTCCACCACTGCTGTTGGCCACACCGCAGACTGCGCTGACT 420
Qy 486 CCCATCCCTCTGAT 500
Db 421 CCCATCCCTCTGAT 435

RESULT 6
AX000970 783 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 15 from Patent WO902696.
DEFINITION AX000970
ACCESSION AX000970
VERSION AX000970.1 GI:7241212
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 783)
AUTHORS Beseme,F. and Blond,J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 902696-A 15 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source 1..783
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 173 a 213 c 166 g 180 t 51 others
ORIGIN

Query Match 82.0%; Score 410; DB 6; Length 783;
Best Local Similarity 93.8%; Pred. No. 6.8e-125;
Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;

Qy 68 TGAAGACAGAGCTAGCTGATTTCTTGAAGCCCACTAAGATCCCTAAGCTGCTGGA 127
Db 1 TGAAGACAGAGCTAGCTGATTTCTTGAAGCCCACTAAGATCCCTAAGCTGCTGGA 60
Qy 128 AGGTGACCACTGCTCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCATCA 187
Db 61 AGGTGACCACTGCTCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCATCA 120
Qy 188 GAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGTAAAGAAATAGCCATCATCT 247
Db 121 GAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGTAAAGAAATAGCCATCATCT 180
Qy 248 ATTGCTGAGAGCAGACAGAGGAGCAACAATCGGATATTAACCCAGGATTCGAGCT 307
Db 181 ATTGCTGAGAGCAGACAGAGGAGCAACAATCGGATATTAACCCAGGATTCGAGCT 240
Qy 308 GGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGATGGAG--CTGTTTCAATGCTA 365
Db 241 GGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCAATGCTA 300
Qy 366 TTTCACTTAATTAATCTTGCAACTGCACTCTGTCGCAATGTTTCTTAAGGCTGAGC 425
Db 301 TTTCACTTAATTAATCTTGCAACTGCACTCTTCTGTCATGTTTCTTAAGGCTGAGC 360
Qy 426 TGAAGCTTTGCTCACCGCTCCACCACTGCTGTTGGCCACACCGCAGACTGCGCTGACT 485
Db 361 TGAAGCTTTGCTCACCGCTCCACCACTGCTGTTGGCCACACCGCAGACTGCGCTGACT 420
Qy 486 CCCATCCCTCTGAT 500
Db 421 CCCATCCCTCTGAT 435

RESULT 7
AX007998 711 bp DNA linear PAT 06-SEP-2000
LOCUS AX007998

DEFINITION Sequence 21 from Patent WO967395.
 ACCESSION AX007998
 VERSION AX007998.1 GI:9995695
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 1 (bases 1 to 711)
 REFERENCE Perin,J.P., Rieger,F. and Alliel,P.M.
 AUTHORS Nucleic sequence and deduced protein sequence family with human
 TITLE endogenous retroviral motifs, and their uses
 JOURNAL Patent: WO 9967395-A 21 29-DEC-1999;
 INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER FRANCOIS (FR); ALLIEL PATRICK M (FR)
 FEATURES
 source
 1. 711
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 171 a 209 c 160 g 171 t
 ORIGIN
 Query Match 76.8%; Score 383.8; DB 6; Length 711;
 Best Local Similarity 95.4%; Pred. No. 3.6e-116;
 Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAGAGACGAGACTGCTGATTTCTAGCGGACTAAGATCCCTAAGCTAGCTGG 125
 |||||
 Db 1 ACTGAGAGACGAGACTGCTGATTTCTAGCGGACTAAGATCCCTAAGCTAGCTGG 60
 QY 126 GAAAGTGACACGCTCACTGCTTAAACAGGGGCTTGAACCTAGCTCACCTGACCAAT 185
 |||||
 Db 61 GAAAGTGACACATCCACTTTAAACAGGGGCTTGAACCTAGCTCACCTGACCAAT 120
 QY 186 CAGAGAGCTCACTAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCCAATCAT 245
 |||||
 Db 121 CAGAGAGCTCACTAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGCCAATCAT 180
 QY 246 CTATTGCTGAGAGACAGGAGGACCAATCGGGATATAACCCAGGCAATTCAG 305
 |||||
 Db 181 CTATTGCTGAGAGACAGGAGGACCAATCGGGATATAACCCAGGCAATTCAG 240
 QY 306 CTGGCAACAGAGCCCCCTTTGGGTCCTTCCCTTTATAGGAG--CTGTTTTCATGC 363
 |||||
 Db 241 CCGGCAACAGGCA--ACCCCTTTGGGTCCCTCCCTTTATAGGAGCTCTGTTTCAATGC 299
 QY 364 TATTCACTATTAAATCTTGCAACTGCACTCTTCTGTCATGTTTCTTACGGCTCGA 423
 |||||
 Db 300 TATTCACTATTAAATCTTGCAACTGCACTCTTCTGTCATGTTTCTTACGGCTCGA 359
 QY 424 GCTAGAGCTTTGCTCAGCGTCCACAGCTGTTTGCACGACGCGCAACTGCGGCTGA 483
 |||||
 Db 360 GCTAGAGCTTTGCTCAGCGTCCACAGCTGTTTGCAGCCGACGCGGCTGA 419
 QY 484 CTCCATCCCTCTGGAT 500
 |||||
 Db 420 CTCCATCCCTCTGGAT 436

RESULT 8
 AL161721/c 137492 bp DNA linear HTG 10-JUL-2001
 LOCUS AL161721
 DEFINITION Homo sapiens chromosome 6 clone RP1-301F24 map q25.2-26, ***
 SEQUENCING IN PROGRESS ***, 6 unordered pieces.
 ACCESSION AL161721
 VERSION AL161721.3 GI:12330750
 KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 REFERENCE 1 (bases 1 to 137492)
 AUTHORS Sims,S.

TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: hunquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jan 22, 2001 this sequence version replaced gi:9212839.
 COMMENT
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: hunquery@sanger.ac.uk
 ----- Project Information
 Center project name: DJ301F24
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 135573 bases at least Q40
 Consensus quality: 135245 bases at least Q30
 Consensus quality: 136166 bases at least Q20
 Insert size: 136992; sum-of-contigs
 Insert size: 152637; 0.5% error; agarose-fp
 Quality coverage: 3.96x in Q20 bases; sum-of-contigs Quality
 coverage: 3.56x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 12245: contig of 12245 bp in length
 * 12246 12345: gap of 100 bp
 * 12346 20428: contig of 8083 bp in length
 * 20429 20528: gap of 100 bp
 * 20529 85104: contig of 64576 bp in length
 * 85105 85204: gap of 100 bp
 * 85205 96812: contig of 11608 bp in length
 * 96813 96912: gap of 100 bp
 * 96913 112622: contig of 15710 bp in length
 * 112623 112722: gap of 100 bp
 * 112723 137492: contig of 24770 bp in length.

FEATURES
 source
 1. 137492
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="q25.2-26"
 /clone="RP1-301F24"
 /clone_1fb="RP1-1"
 1. 12245
 /note="assembly_fragment:00933
 fragment_chain:1
 clone_end:17
 vector_side:left"
 12346_20428
 /note="assembly_fragment:01071
 fragment_chain:1"
 20529_85104
 /note="assembly_fragment:00093
 fragment_chain:1"
 85205_96812
 /note="assembly_fragment:00954
 fragment_chain:1"
 96913_112622
 /note="assembly_fragment:01341"
 112723_137492
 /note="assembly_fragment:00790
 clone_end:SP6
 vector_side:right"
 BASE COUNT 37669 a 28026 c 30367 g 40929 t 501 others
 ORIGIN

Query Match 72.0%; Score 360; LB 2; Length 137492;
Best Local Similarity 89.6%; Pred. No. 4.3e-108;
Matches 398; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

Qy 57 TATCTGCACTGAGACAGACGACTGAGTTTCCTTAGCCGACTAAGAAATCCCTTAAG 116
Db 128053 TTTAATGGTCTTGAGAGACAGACTAGCTGATTCCTTAGGCCGACTAAGAAATCCCTTAAG 127994
Qy 117 CCTACTGGGAAGGTGACACGCTCCACTTAACAGGGGCTTGCACTTAGCTCACAC 176
Db 127993 CCTACTGGGAAGGTGACACGCTCCACTTAACAGGGGCTTGCACTTAGCTCACAC 127934
Qy 177 CTGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGTAAAGAAATA 236
Db 127993 CCAACCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGTAAAGAAATA 127874
Qy 237 GCCAATCATCTATTGCTGAGACACAGAGAGGAACAACATGGGATATTAACCCAG 296
Db 127873 GCCAATCATCTATTGCTGAGACACAGAGAGGAACAACATGGGATATTAACCCAG 127814
Qy 297 GCATCGAGCTGGCAACAGACGCCCCCTTGGGTCCCTTCCCTTGTATGGAGCTGT 356
Db 127813 GCATCGAGCTGGCAACAGACGCCCCCTTGGGTCCCTTCCCTTGTATGGAGCTGT 127755
Qy 357 TTCATGCTAATTTCACTCTATTAAATCTTGCACTGCATCTTCTGTGCTCCATGTTCTTAC 416
Db 127754 TTCATGCTAATTTCACTCTATTAAATCTTGCACTGCATCTTCTGTGCTCCATGTTCTTAC 127695
Qy 417 GGCTGAGCTGAGCTTTTGTCTGACCGTCCACCACTGCCTTTTGCAACACCGCAGACCTG 476
Db 127694 GGCTGAGCTGAGCTTTTGTCTGACCGTCCACCACTGCCTTTTGCAACACCGCAGACCTG 127635
Qy 477 CCGCTGACTCCCATCCCTCTCGAT 500
Db 127634 CCGCTGACTCCCATCCCTCTCGAT 127611

RESULT 9
HS836J3/C
LOCUS HS836J3 143590 bp ENA linear PRI 02-MAR-2000
DEFINITION Human DNA sequence from clone RP5-836J3 on chromosome 1p22.1-31.1
ACCESSION AL035706
VERSION AL035706.10 GI:5777502
KEYWORDS HTG; Cpg island.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Crania-ia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 143590)
AUTHORS Bagunley C.
TITLE Direct Submission
JOURNAL Submitted (29-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humguery@sanger.ac.uk
COMMENT This sequence has been finished accoridng to sequence map criteria
as follows. An attempt is made to remove all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/chr1
RP5-836J3 is from the library RPc1-5 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP5-836J3.

FEATURES

source
1..143590
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="P22.1-31.1"
/clone="RP5-836J3"
/clone_1lb="RPc1-5"
1160..1786
/note="match: GSS: Em:B16380"
complement(2068..2500)
/note="match: GSS: Em:AQ089628"
2554..3072
/note="match: GSS: Em:AQ118130"
complement(3824..4133)
/note="match: GSS: Em:AQ089896"
4017..4402
/note="match: GSS: Em:AQ063212"
4017..4312
/note="match: GSS: Em:AQ058070"
4023..4278
/note="match: GSS: Em:AQ063209"
6000..6045
/note="2 copies 23 mer 100% conserved"
6337..6906
/note="ONSU repeat: matches 190..467 of consensus"
9859..10375
/note="MLT1F repeat: matches 3..536 of consensus"
13112..13891
/note="LTR17 repeat: matches 1..780 of consensus"
15130..15173
/note="22 copies 2 mer tg 93% conserved"
15133..15172
/note="4 copies 10 mer gtgtgtgtgt 95% conserved"
23083..23118
/note="18 copies 2 mer tg 80% conserved"
23328..23713
/note="WSTD repeat: matches 1..394 of consensus"
26442..26509
/note="34 copies 2 mer gt 66% conserved"
complement(26847..27324)
/note="match: GSS: Em:AQ835456"
27344..27699
/note="match: GSS: Em:AQ479571"
31575..32054
/note="match: STS: Em:AL009625"
32094..32491
/note="MLT2D repeat: matches 1..490 of consensus"
32492..32857
/note="match: GSS: Em:AQ140889"
32521..32580
/note="6 copies 10 mer gatagataga 80% conserved"
32587..32587
/note="15 copies 4 mer atag 96% conserved"
32591..32639
/note="MLT2D repeat: matches 505..553 of consensus"
38429..38876
/note="match: GSS: Em:AQ146760"
39727..39756
/note="6 copies 5 mer tggct 96% conserved"
40414..40453
/note="20 copies 2 mer ca 97% conserved"
42939..42989
repeat_region

/note="MER91A repeat: matches 95. .152 of consensus"
43515. .43610
/note="2 copies 48 mer 85% conserved"
misc_feature
/note="match: GSS: Em:AQ437799"
44168. .44654
/note="match: GSS: Em:AQ564589"
45309. .45340
/note="16 copies 2 mer tt 93% conserved"
misc_feature
45341. .45660
/note="match: STS: Em:G24052"
45473. .45610
/note="6 copies 23 mer 67% conserved"
repeat_region
46017. .46066
/note="25 copies 2 mer ac 78% conserved"
repeat_region
46019. .46066
/note="12 copies 4 mer acac 79% conserved"
repeat_region
49244. .49451
/note="MER58A repeat: matches 1. .223 of consensus"
51847. .51898
/note="13 copies 4 mer ctct 73% conserved"
repeat_region
51849. .51898
/note="25 copies 2 mer ct 74% conserved"
misc_feature
/note="match: GSS: Em:AQ440464"
56014. .56544
/note="MER68B repeat: matches 1. .568 of consensus"
57696. .57814
/note="MER69 repeat: matches 2. .129 of consensus"
58171. .58704
/note="MLT1E repeat: matches 3. .548 of consensus"
66786. .67150
/note="match: GSS: Em:AQ053163"
67192. .67597
/note="MER7A repeat: matches 1. .346 of consensus"
/note="complement (68816. .69390)
/note="match: GSS: Em:AQ549150"
69397. .69432
/note="9 copies 4 mer taag 86% conserved"
misc_feature
69519. .70015
/note="match: GSS: Em:AQ045852"
69911. .69986
/note="19 copies 4 mer taat 65% conserved"
repeat_region
70512. .71004
/note="MER4A repeat: matches 110. .656 of consensus"
71303. .71412
/note="MER4A repeat: matches 1. .110 of consensus"
/note="complement (73648. .74145)
/note="match: GSS: Em:AQ458708"
75280. .75333
/note="27 copies 2 mer tt 72% conserved"
repeat_region
76040. .76115
/note="19 copies 4 mer aaga 97% conserved"
repeat_region
76046. .76115
/note="35 copies 2 mer ga 75% conserved"
repeat_region
76047. .76126
/note="8 copies 10 mer aagaaga 76% conserved"
repeat_region
76923. .79419
/note="MLT1F repeat: matches 67. .541 of consensus"
80005. .80032
/note="7 copies 4 mer atac 96% conserved"
repeat_region
82920. .83118
/note="MER91A repeat: matches 2. .195 of consensus"
83123. .83171
/note="MER91A repeat: matches 138. .186 of consensus"
83659. .84165
/note="L1PB2 repeat: matches 5634. .6155 of consensus"
87050. .87201
/note="MER4A repeat: matches 1. .148 of consensus"
87197. .87536
/note="MER4A repeat: matches 394. .728 of consensus"
91670. .92092
/note="match: GSS: Em:AQ275888"

misc_feature 91670. .92091
/note="match: GSS: Em:AQ279665"
misc_feature 91670. .92093
/note="match: GSS: Em:AQ278706"
repeat_region 93141. .94747
/note="MST4-internal repeat: matches 1. .1637 of consensus"
repeat_region 94761. .95183
/note="MSTB repeat: matches 1. .426 of consensus"
repeat_region 95891. .96457
/note="L1MCI repeat: matches 5219. .5796 of consensus"
repeat_region 96486. .96839
/note="MER74A repeat: matches 165. .524 of consensus"
repeat_region 96940. .97505
/note="L1MCI repeat: matches 5791. .6325 of consensus"
repeat_region 100936. .100997
/note="1us repeat: matches 230. .291 of consensus"
repeat_region 105831. .106199
/note="LOR1B repeat: matches 1. .461 of consensus"
repeat_region 106592. .107055
/note="L1MBS repeat: matches 5688. .6174 of consensus"
repeat_region 107066. .107248

Query Match 72.0%; Score 359.8; DB 9; Length 143590;
Best Local Similarity 88.3%; Pred.No. 4.8e-108;
Matches 414; Conservative 0; Mismatches 52; Indels 3; Gaps 2;

QY AAAAGCGCTGGAGATTCAGCAATTCCTGCACTGAGAGACGAGCTGAGATTTC 93
DB 13925 AAAATGATAGCAGCTCCAGTATTTAGAGACATCTGAGATAGACTGATTTCC 13866
QY TAGCGCGACTAGAAATCCCTTAAGCTGAGAGAGTGCACGTCACCTTTAAACAC 153
DB 13865 TAGCGCGACTAGAAATCCCTTAAGCTGAGAGAGTGCACGTCACCTTTAAACAT 13806
QY 154 GGGCGTTGCACTTACTGACACCTGACCAATAGAGAGCTCAATTAATTTAGG 213
DB 13805 GGGCGTTGCACTTACTGACACCTGACCAATAGAGAGCTCAATTAATTTAGG 13746
QY 214 CAAAGCAGAGAGTAAAGAAATAGCAATCATCTATTTGCTGAGAGACGAGAGGGA 273
DB 13745 CAAAGCAGAGAGTAAAGAAATAGCAATCATCTATTTGCTGAGAGACGAGAGGGA 13686
QY 274 CAAATTCGGGATATAAACCCAGGAGCTGAGCAAGAGCCCTTTGGGTC 333
DB 13685 CAAATTCGGGATATAAACCCAGGAGCTGAGCAAGAGCCCTTTGGGTC 13627
QY 334 CTTCCTTTGATGAGAG-CTGTTTCATGCTATTCTACTATTAATCTTGCAACTG 391
DB 13626 CTTCCTTTGATGAGAGCTGTTTCACTCTATTAAATCTTGCAACTG 13567
QY 392 CACTCTTCGTCGATGTTTCTTAACGCTGAGCTGAGCTTTGCTCACCGTCCACCACT 451
DB 13566 CACTCTTCGTCGATGTTTCTTAACGCTGAGCTTTGCTGCGTCCACCACT 13507
QY 452 GCTGTTTGCCACGACGAGACCTGCGCTGACTCCATCCCTTGAGAT 500
DB 13506 GCTGTTTGCCACGACGAGACCTGCGCTGACTCCATCCCTTGAGAT 13458

RESULT 10
AC103595/c 119406 bp DNA linear HTG 29-NOV-2001
LOCUS Homo sapiens chromosome 1 clone RP11-415P10, WORKING DRAFT
DEFINITION SEQUENCE, 10 unordered pieces.
AC103595
AC103595.1 GI:7149460
VERSION HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 119406)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.

TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 119406)
AUTHORS	Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE	Direct Submission
JOURNAL	Submitted (29-NOV-2001) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT	----- Genome Center Center: University of Washington Genome Center Center Code: UMG Web site: http://www.genome.washington.edu Contact: uwgchgs@u.washington.edu ----- Project Information Center project name: chr-1 Center clone name: RP11-415P10 (sc0362) ----- Summary Statistics Sequencing vector: Plasmid; 108752; 100% of reads Chemistry: Dye-terminator ET; 92% of reads Chemistry: Dye-terminator Big dye; 8% of reads Assembly program: Phrap; version 0.99019 Consensus quality: 109646 bases at least Q40 Consensus quality: 115473 bases at least Q30 Consensus quality: 117774 bases at least Q20 Insert size: 194800; 40.1% error; agarose-fp Insert size: 118506; sum-of-contigs Quality coverage: 5.7x in Q20 bases; agarose-fp Quality coverage: 9.3x in Q20 bases; sum-of-contigs ----- * NOTE: This is a 'working draft' sequence. It currently * consists of 10 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. * * 1 4564: contig of 4564 bp in length * 4565 4664: gap of unknown length * 4665 12930: contig of 8266 bp in length * 12931 13030: gap of unknown length * 13031 20784: contig of 7754 bp in length * 20785 20884: gap of unknown length * 20885 29093: contig of 8209 bp in length * 29094 29193: gap of unknown length * 29194 40511: contig of 1116 bp in length * 40512 40611: gap of unknown length * 40612 52409: contig of 11798 bp in length * 52410 52509: gap of unknown length * 52510 64775: contig of 12266 bp in length * 64776 64875: gap of unknown length * 64876 83398: contig of 18523 bp in length * 83399 83499: gap of unknown length * 83499 100056: contig of 16558 bp in length * 100057 100156: gap of unknown length * 100157 119406: contig of 19250 bp in length. Location/Qualifiers 1. 119406 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="1" /clone="RP11-415P10" /clone_1b="PPCI human BAC library 11" 1. 4564 /note="assembly_name:Contig:9" 4665 .12930 /note="assembly_name:Contig:0" 13031 .20784 /note="assembly_name:Contig:1" 20885 .29093 /note="assembly_name:Contig:2" 29194 .40511 /note="assembly_name:Contig:3" 40612 .52409

	misc_feature	/note="assembly_name:Contig34"	52510..64775	
	misc_feature	/note="assembly_name:Contig35"	64876..83398	
	misc_feature	/note="assembly_name:Contig36"	83499..100056	
	misc_feature	/note="assembly_name:Contig37"	100157..119406	
	BASE COUNT	35588 a 23913 c 23787 g 35137 t	981 others	
	ORIGIN			
	Query Match	71.6%; Score 358.2; DB 2;	Length 119406;	
	Best Local Similarity	88.1%; Pred. No. 1.6e-107;		
	Matches 413; Conservative 0; Mismatches 53; Indels 3; Gaps 2;			
Oy	34	AAAGCCTGAGATACGCAATTATCTTGCACTGAGACAGACACTGAGTTCC	93	
Db	37607	AAAATGATAGCGCTACAGATTAGAGACATCTGAGATAGAGACTGAGATTCC	37548	
Oy	94	TAGGCCGACTAAGAATTCCTTAAGCTTAGCTGGGAAGTGAACAGTCCATTAAACAC	153	
Db	37547	TAGGCCGACTAAAAATCCCTAAGCCTAGTTGGGAAGTGACCGCATCACCTTAAACAT	37488	
Oy	154	GGGGCTTGCACTTAGCTGCACACCTGACCATTCAGAGAGCTCAGTAAATGTATTAGG	213	
Db	37487	GGGGCTTGCACTTAGCTGCACACCTGACCATTCAGAGAGCTCAGTAAATGTATTAGG	37428	
Oy	214	CAGAAGCAGAGGTAAAGAAATAGCCAATCATCTATTGCTTGAGAGCAGACGAGGGA	273	
Db	37427	CAGAAACAGAGGTAGAGAAATAGCCAATCATCTATTGCTTGAGAGCAGAGTGGGA	37368	
Oy	274	CACAATATGGGATATAAACCCAGGATTCGAGCTGGGCAACGACCCCCCTTTGGGTC	333	
Db	37367	CAGAGATGGGATATAAACCCAGGATTCGAGCTGGGCAACGACCCCCCTTTGGGTC	37309	
Oy	334	CTTCCCTTTGATGGGAG--CTGTTTCATGCTATTCTACTCTATTAACTTTGCACTG	391	
Db	37308	CTTCCCTTTGATGGGAGCTCTGTTTACCTATTTTACCTATTAATCTTTGCACTG	37249	
Oy	392	CACCTCTGATGCTCATGTTTTCTTACGCTGAGAGCTTTTGCTCAACGTCACACT	451	
Db	37248	CACCTCTGATGCTCATGTTTTCTTACGCTGAGAGAGCTTTTGCTTCCGTCACACT	37189	
Oy	452	GCTGTTTGGCACACCGAGACCTGCGCTGACTCCATCCCTCTCGAT	500	
Db	37188	GCTGTTTGGCACACCGAGACCTGCGCTGACTCCATCCCTCTCGAT	37140	
	RESULT 11			
	ALJ53584_0/c			
	WPCOMMENT			
	Sequence split into 6 fragments	LOCUS ALJ53584	Accession ALJ53584	
	Fragment Name	Begin	End	
	ALJ53584_0	1	110000	
	ALJ53584_1	100001	210000	
	ALJ53584_2	200001	310000	
	ALJ53584_3	300001	410000	
	ALJ53584_4	400001	510000	
	ALJ53584_5	500001	536214	
	LOCUS	ALJ53584	536214 bp	DNA linear HTG 19-JUL-2001
	DEFINITION	Homo sapiens chromosome X clone RP13-185F3,	*** SEQUENCING IN	
	PROGRES ***	in ordered pieces.		
	ACCESSION	ALJ53584		
	VERSION	ALJ53584.11	GI:15020381	
	KEYWORDS	HTG; HTGS_PHASE2; HTGS_CANCELLED.		
	SOURCE	human.		
	ORGANISM	Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	REFERENCE	1 (bases 1 to 536214)		
	AUTHORS	Wilson, S.		
	TITLE	Direct Submission		

Query Match 71.6%; Score 357.8; DB 9; Length 169029;
 Best Local Similarity 92.0%; Pred. No. 2.3e-107;
 Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;

QY 68 TGAGAGACAGAGACTAGCTGATTTCTTAGGCCGACTAGAAATCCCTTAAGCTTAGCTGGGA 127
 DB 22071 TGAGAGACAGAGACTAGCTGATTTCTTAGGCCGACTAGAAATCCCTTAAGCTTAGCTGGGA 22012

QY 128 AGGTGACACAGCTCCACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAATCA 187
 DB 22011 AGGTGACACAGCTCCACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAATCA 21952

QY 188 GAGAGCTCACTAAATCTAATTAGGCAAGACAGGA 3GTAAAGAAATAGCCAATCATCT 247
 DB 21951 GAGAGCTCACTAAATCTAATTAGGCAAGAAACAGGA 3GTAAAGAAATAGCCAATCATCT 21892

QY 248 ATTGCTGAGAGCAGACAGAGAGGAGCAACAATGGGTTATTAACCAAGCATTCGAGCT 307
 DB 21891 ATTGCTGAGAGCAGACAGAGAGGAGCAACAATGGGTTATTAACCAAGCATTCGAGCC 21832

QY 308 GGGACAGACAGCCCCCTTTGGGTCCTTCCTTTGTGTGGGAG--CTGTTTTCATCTA 365
 DB 21831 GGGACAGTGA-ACCCCTTTGGGTCCTTCCTTTGTGTGGGAGCTCTGTTTTCATCTA 21773

QY 366 TTTCACCTAATTAAATCTTGCACTGCACTCTTCTGG:CCATGTTTCTTAAGGCTCGAGC 425
 DB 21772 TTTCACCTAATTAAATCTTGCACTGCACTCTTCTGG:CCATGTTTCTTAAGGCTCGAGC 21713

QY 426 TGAGCTTTTGTCAACGCTCCACCACTGCTGTTTGGCA(CACCGAGACCTGCGCTGACT 485
 DB 21712 TGAGCTTTTGTCTCCCACTCCACCACTGCTGTTTGGCG(CATTAGACAGCCGCGCTGACT 21653

QY 486 CCCATCCCTCGGAT 500
 DB 21652 CCCATCCCTCGGAT 21638

RESULT 13
 AL392173/c 170746 bp DNA linear PRI 05-APR-2001
 LOCUS DEFINITION Human DNA sequence from clone RP11-158M9 on chromosome Xq26.1-27.1
 AL392173
 AL392173.8 GI:11064224
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 170746)
 Bray-Allen,S.
 Direct Submission
 Submitted (07-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Oct 31, 2000 this sequence version replaced gi:10800700.
 During sequence assembly data is compared from overlapping clones.
 where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome X, constructed by the Sanger Centre Chromosome X Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
 This sequence is the entire insert of clone RP11-158M9 This
 sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate
 Chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP11-158M9 is from
 the library RPCR-11.1 constructed by the group of Pieter de Jong.
<http://www.ncbi.org/bacpac/home.htm>
 VECTOR: pBAC3.6.

FEATURES

source

1..170746
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /map="q26.1-27.1"
 /clone="RP11-158M9"
 /clone_11b="RPCR-11.1"
 688..1038
 /note="match: GSS: Em:AQ314307"
 2318..2343
 /note="13 copies 2 mer ca 100% conserved"
 7251..7937
 /note="11P6 repeat: matches 4764..5447 of consensus"
 7938..8568
 /note="11P6 repeat: matches 5507..6139 of consensus"
 9828..10340
 /note="match: GSS: Em:A2368366"
 14268..14619
 /note="MR34 repeat: matches 4..301 of consensus"
 20566..21107
 /note="MLT2B repeat: matches 1..448 of consensus"
 27734..27963
 /note="match: STS: Em:HS210VD10"
 27963..27998
 /note="18 copies 2 mer ac 100% conserved"
 27967..28062
 /note="match: STS: Em:HS210VD10"
 29226..29569
 /note="THEIC repeat: matches 1..371 of consensus"
 29570..30022
 /note="THEIC-internal repeat: matches 1..445 of consensus"
 30335..31434
 /note="THEIC-internal repeat: matches 445..1580 of consensus"
 31435..31807
 /note="THEIC repeat: matches 1..371 of consensus"
 35205..35669
 /note="match: GSS: Em:AQ493428"
 37464..38246
 /note="LTR17 repeat: matches 1..780 of consensus"
 39054..39117
 /note="13 copies 2 mer gc 92% conserved"
 41432..41546
 /note="LTR29 repeat: matches 516..619 of consensus"
 41877..42007
 /note="LTR29 repeat: matches 3..149 of consensus"
 51113..51140
 /note="14 copies 2 mer tc 89% conserved"
 51268..51318
 /note="HY5 repeat: matches 30..76 of consensus"
 54978..55013
 /note="18 copies 2 mer ca 91% conserved"
 58352..59188
 /note="THE1B-INTERNAL repeat: matches 1..842 of consensus"
 61106..61272
 /note="WSTB-internal repeat: matches 15..186 of consensus"
 61578..61999
 /note="WSTB-internal repeat: matches 186..585 of consensus"
 62003..62098
 /note="WST-INTERNAL repeat: matches 594..689 of consensus"
 63045..63698

```
/note="HERV16 repeat: matches 4. .664 of consensus"
repeat_region 63927. .64774
/note="HERV16 repeat: matches 716. .1577 of consensus"
repeat_region 64789. .64832
/note="22 copies 2 mer aa 81% conserved"
repeat_region 64916. .65612
/note="HERV16 repeat: matches 1817. .2518 of consensus"
repeat_region 65613. .66886
/note="THE1-INTERNAL repeat: matches 124. .1401 of consensus"
misc_feature 66475. .66484
/note="1372 bp of 18186 transposon (X03123) removed here. This sequence represents the duplicated flanking sequence of the 18186."
70601. .70772
/note="match: GSS: Em:AQ077330"
70734. .71229
/note="match: GSS: Em:AQ628837"
80000. .80512
/note="match: GSS: Em:AQ284023
match: STS: Em:G61704"
complement(81135. .81789)
/note="match: GSS: Em:AQ240981"
81193. .81655
/note="match: GSS: Em:AQ704174"
83625. .89762
/note="LTPA5 repeat: matches 11. .6145 of consensus"
91802. .91895
/note="LTPA3 repeat: matches 6053. .6146 of consensus"
94922. .101052
/note="LTPA7 repeat: matches 11. .6145 of consensus"
102779. .102828
/note="25 copies 2 mer ac 82% conserved"
complement(103577. .104157)
/note="match: GSS: Em:AQ775154"
105177. .105697
/note="match: GSS: Em:B63384"
105842. .106341
/note="match: GSS: Em:AQ522879"
107797. .108270
/note="MER39 repeat: matches 13. .521 of consensus"
116826. .116918
/note="MER91A repeat: matches 1. .94 of consensus"
complement(122839. .123304)
/note="match: GSS: Em:AQ002183"
126653. .126696
/note="22 copies 2 mer ac 93% conserved"
repeat_region 127908. .127950
/note="LTPA5 repeat: matches 6103. .6145 of consensus"
complement(129916. .130692)
/note="match: GSS: Em:A2288149"
complement(130168. .130745)
/note="match: GSS: Em:A2293756"
complement(131672. .132168)
/note="match: GSS: Em:AQ466443"
131821. .131846
/note="13 copies 2 mer ta 96% conserved"
132163. .132600
/note="match: GSS: Em:AQ829322"
133493. .133664
/note="LTPA3 repeat: matches 30. .202 of consensus"
136095. .136245
/note="LTPA7 repeat: matches 6142. .6289 of consensus"
complement(140979. .141356)
/note="match: STS: Em:HSC28H08"
146003. .146304
/note="LTPA3 repeat: matches 1. .310 of consensus"
146351. .146421
/note="LTPA3 repeat: matches 448. .516 of consensus"
157021. .157123
/note="MER34 repeat: matches 76. .173 of consensus"
159023. .159689
/note="HERV16 repeat: matches 562. .1217 of consensus"
```

```
repeat_region 159978. .160033
/note="HERV16 repeat: matches 1217. .1272 of consensus"
repeat_region 164301. .166134
/note="LTPA14 repeat: matches 1841. .3688 of consensus"
repeat_region 166458. .166509
/note="26 copies 2 mer ga 73% conserved"
repeat_region 166568. .169037
/note="LTPA14 repeat: matches 3681. .6148 of consensus"
misc_feature 169389. .169794
/note="match: GSS: Em:AQ092454"
misc_feature 170139. .170746
complement(170286. .170416)
/note="match: GSS: Em:AQ389950"
misc_feature 170286. .170416
complement(170286. .170416)
/note="match: GSS: Em:AQ835692"
BASE COUNT 53940 a 32801 c 34298 g 49707 t
ORIGIN
Query Match 71.6%; Score 357.8; DB 9; Length 170746;
Best Local Similarity 92.0%; Pred. No. 2.3e-107;
Matches 400; Conservative 0; Mismatches 32; Indels 3; Gaps 2;
QY 68 TGAGAGCAGGACTAGCTGATTTCTTAGGCGCACTAAGATCCCTAAGCTTAGCTGGGA 127
Db 38246 TGAGAGCAGGACTAGCTGATTTCTTAGGCGCACTAAGATCCCGAAGCTTAGCTGGGA 38187
QY 128 AGGTGACAGGCTGACCTTTAAACAGGGGCTTGCACTTAGCTCACCGTGAACATCA 187
Db 38186 AGGTGACAGGCTGACCTTTAAACAGGGGCTTGCACTTAGCTCACCGTGAACATCA 38127
QY 188 GAGAGCTCACTAAATGCTATTAGGCAAGAGAGTAAAGAAATAGCAATCATCT 247
Db 38126 GAGAGCTCACTAAATGCTATTAGGCAAGAGAGTAAAGAAATAGCAATCATCT 38067
QY 248 ATTGCTGAGAGCAGCAGAGGAGACAAATCGGATTTAAACCGAGCTTCAGCT 307
Db 38066 ATTGCTGAGAGCAGCAGTGGAGAGGAAAGATCGGATTTAAACCGAGCTTCAGCT 38007
QY 308 GGAACAGAGCCCGCTTTGGGTCCTTCCTTTGATGGAG--CTGTTTCATGCTA 365
Db 38006 GGAACAGGCA-ACCCCTTTGGGTCCTTCCTTTGATGGAGCTGTTTCATGCTG 37948
QY 366 TTTCACCTATTAAATCTTGCAACTGCACTCTTGCTGATGTTCTTAAGGCTCGAGC 425
Db 37947 TTTCACCTATTAAATCTTGCAACTGCACTCTTGCTGATGTTCTTAAGGCTCGAGC 37888
QY 426 TGAGCTTTGCTCAGCGTCCACCACTGCTGTTGCCACCGAGACTCGGCTACT 465
Db 37887 TGAGCTTTGCTCAGCGTCCACCACTGCTGTTGCCCGCGAGACCGCGCTGACT 37828
QY 486 CCCATCCCTCGGAT 500
Db 37827 TCCATCCCTCGGAT 37813
RESULT 14
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS AC093531 Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
DEFINITION AC093531
VERSION AC093531.2 GI:16945981
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163803)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
```

REFERENCE 3 (bases 1 to 163803)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Nov 16, 2001 this sequence version replaced gi:15383820.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 finishing completed at Stanford Human Genome Center
 www-shgc.stanford.edu
 Quality: Phrap Quality >=40 99.8% cf Sequence;
 Estimated Total Number of Errors is 0.3.
 Location/Qualifiers
 1. 163803
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="RP11-4051.7"

BASE COUNT 54797 a 31372 c 29086 g 48548 t
 ORIGIN

Query Match 71.5%; Score 357.4; DB 9; Length 163803;
 Best Local Similarity 91.2%; Pred. No. 3.1e-107;
 Matches 402; Conservative 0; Mismatches 36; Indels 3; Gaps 2;

QY 62 TGCACCTAGAGACAGCACTGCTGATTTCTTAGCGCTGACCTAGAGAAATCCCTAGCCTT 121
 DB 120271 TACCAATTAGACACAGCACTGCTGATTTCTTAGCGCTGACCTAGAGAAATCCCTAGCCTT 120212
 QY 122 CTGGGAGGTGACCACTGCTGATTTCTTAGCGCTGACCTAGAGAAATCCCTAGCCTT 181
 DB 120211 CTGGGAGGTGACCTGCTGATTTCTTAGCGCTGACCTAGAGAAATCCCTAGCCTT 120152
 QY 182 CAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAAATGCCCCA 241
 DB 120151 CAATCAGAGAGCTCACTAAATGCTAATTAGGCAAGAAACAGAGGTAAGAAATGCCCCA 120092
 QY 242 TCATCTATTGCTGAGAGACAGGAGGAGCAACATTCGGGATATTAACCCAGGCAAT 301
 DB 120091 TCATCTATTGCTGAGAGACAGGAGGAGCAACATTCGGGATATTAACCCAGGCAAT 120032
 QY 302 CGAGCTGACACAGCAGCCCCCTTTGGGTCTCCCTTTGTATGGAG--CTGTTTTC 359
 DB 120031 TGAGCTGGGAATGGCA-ACCCCTTTGGGTCTCCCTTTGTATGGAGCTCTGTTTTC 119973
 QY 360 ATGCTATTCTCACTAATTAATCTTGCAATGCACTCTCTGCTGCTATGTTCTTAAGGC 419
 DB 119972 ACTCTATTCTCACTAATTAATCTTGCAATGCACTCTCTGCTGCTATGTTCTTAAGGC 119913
 QY 420 TCGAGCTAGAGCTTTTGTACAGCGTCACAGCTGCTTTGACCAACCGCAGACCTGCGC 479
 DB 119912 TCAAGCTAGAGCTTTTGTACAGCGTCACAGCTGCTTTGACCAACATTCGGAAGACCACTG 119853
 QY 480 CTGACTCCCATCCCTCTGGAT 500
 DB 119852 CTGACTTCATCCCTCCAGAT 119832

RESULT 15
 AC010888
 LOCUS Homo sapiens 191863 bp DNA linear PRI 07-NOV-2001
 DEFINITION AC010888 BAC clone RP11-418N16 from 2, complete sequence.
 ACCESSION AC010888
 VERSION AC010888.12 GI:13992760
 KEYWORDS HTG.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 191863)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 191863)
 AUTHORS Margulis, C., Abbott, A. and Sawyer, C.
 TITLE The sequence of Homo sapiens BAC clone RP11-418N16
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 191863)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 4 (bases 1 to 191863)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (08-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 5 (bases 1 to 191863)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 6 (bases 1 to 191863)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On May 8, 2001 this sequence version replaced gi:13431186.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: saplens@wustl.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0418N16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 The RP11-418N16 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanesi, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://pacpac.med.buffalo.edu)
 VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-434M17, 200 bp overlap. Actual start of this clone is at base position 195 of RP11-418N16; actual end is at base position 191863 of RP11-418N16.

The sequence RP11-418N16 between bases 117880 to 117930 is covered by a single plasmid subclone that has some low quality regions.

FEATURES

Source

```
1. 191863
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
   /clone="RP11-418N16"
   /clone_11b="RPC1-11"
   387..700
repeat_region
   /rpt_family="Alu"
   906..1517
repeat_region
   /rpt_family="MaLR"
   2319..2372
repeat_region
   /rpt_family="AT_rich"
   2456..3313
   /rpt_family="L1"
   3314..3345
repeat_region
   /rpt_family="(TAA)n"
   3346..4533
repeat_region
   /rpt_family="L1"
   4550..4679
repeat_region
   /rpt_family="CT-rich"
   4680..6289
repeat_region
   /rpt_family="SVA"
   4839..5815
misc_feature
   /note="CpG island (GC=68.7, o/e=0.70, #CpGs=68)"
   6291..6330
repeat_region
   /rpt_family="L1"
   6331..6642
repeat_region
   /rpt_family="Alu"
   6643..7867
repeat_region
   /rpt_family="L1"
   7870..8126
repeat_region
   /rpt_family="Mariner"
   8127..8491
repeat_region
   /rpt_family="L1"
   8493..9037
repeat_region
   /rpt_family="ERVU"
   9038..10102
repeat_region
   /rpt_family="ERVU"
   10097..10959
repeat_region
   /rpt_family="ERVU"
   10960..11329
repeat_region
   /rpt_family="MaLR"
   12986..13361
repeat_region
   /rpt_family="MaLR"
   13362..15300
repeat_region
   /rpt_family="ERVU"
   15300..15696
repeat_region
   /rpt_family="ERVU"
   15720..16137
repeat_region
   /rpt_family="Tc2"
   16149..16206
repeat_region
   /rpt_family="Mariner"
   16210..16313
repeat_region
   /rpt_family="Tc2"
   16571..16706
repeat_region
   /rpt_family="L2"
   17680..17752
repeat_region
   /rpt_family="L1"
   17777..17859
repeat_region
   /rpt_family="MIR"
   18083..18285
repeat_region
   /rpt_family="L1"
   18546..18879
repeat_region
   /rpt_family="L1"
   22173..22474
repeat_region
   /rpt_family="Alu"
   23231..24056
repeat_region
   /rpt_family="MER2_type"
   24453..24476
```

```

   /rpt_family="AT_rich"
   24477..24768
repeat_region
   /rpt_family="L1"
   24769..24801
repeat_region
   /rpt_family="AT_rich"
   24908..25040
repeat_region
   /rpt_family="MIR"
   25445..25761
repeat_region
   /rpt_family="MER1_type?"
   26238..26534
repeat_region
   /rpt_family="Alu"
   26605..26932
repeat_region
   /rpt_family="MER1_type?"
   27139..27264
repeat_region
   /rpt_family="MIR"
   27364..27706
misc_feature
   /note="similar to EST AW692638 (NTD:98056843)"
   27715..27889
repeat_region
   /rpt_family="MIR"
   28246..28385
repeat_region
   /rpt_family="L2"
   28665..28689
repeat_region
   /rpt_family="AT_rich"
   28925..29245
repeat_region
   /rpt_family="Alu"
   29844..30032
repeat_region
   /rpt_family="L1"
   30033..30079
repeat_region
   /rpt_family="US"
   30080..30487
repeat_region
   /rpt_family="L1"
   30488..30650
repeat_region
   /rpt_family="MER1_type"
   30651..30929
repeat_region
   /rpt_family="Alu"
   30930..30974
repeat_region
   /rpt_family="MER1_type"
   30975..31085
repeat_region
   /rpt_family="L1"
   31211..31417
repeat_region
   /rpt_family="MIR"
   31455..31632

Query Match          71.5%; Score 357.4; DB 9; Length 191863;
Best Local Similarity 91.2%; Pred. No. 3,1e-107;
Matches 402; Conservative 0; Mismatches 36; Indels 3; Gaps 2;

QY 62 TGCACCTGAGAGACAGGACTGATGTTTCCTAGGCCGACTAAGATCCCTAAGCCTAG 121
   |||||
Db 171376 TGCACCTGAGAGACAGGACTGATGTTTCCTAAGGCCGACTAAGATCCCTAAGCCTAG 171435

QY 122 CTGGGAAGGTGACCACTCCACTTTTAAACAGGGGCTTGCAACTTGTGCTACACCTGAC 181
   |||||
Db 171436 CTGGGAAGGTGACCACTCCACTTTTAAACAGGGGCTTGCAACTTGTGCTACACCTGAC 171495

QY 182 CAATCAGAGAGCTCATTAAATGCTAATTAGGCAAGAGAGGTAAAGAAATPAGCAA 241
   |||||
Db 171496 CAATCAGAGAGCTCATTAAATGCTAATTAGGCAAGAGAGGTAAAGAAATPAGCAA 171555

QY 242 TCATCTATTGCTGAGAGCAGCAGAGAGGAGACAACATCGGATATPAAACCAAGCATT 301
   |||||
Db 171556 TCATCTATTGCTGAGAGCAGCAGAGAGGAGACAAGATCGGATATPAAACCAAGCATT 171615

QY 302 CGAGCTGGCAACAGCAGCCCCCTTTGGTCCCTTCCTTTGTATGGAG--CTGTTTC 359
   |||||
Db 171616 CGAGCTGGCAACAGCAGCCCCCTTTGGTCCCTTCCTTTGTATGGAGCTGTTTC 171674

QY 360 ATGTATTCTACTCTATTAATCTTGCACTGCACTCTGCGCATGTTCTTAAGGC 419
   |||||
Db 171675 ACTTATTCTACTCTATTAATCTTGCACTGCACTCTGCGCATGTTCTTAAGGC 171734

QY 420 TCGAGTGAAGCTTTGCTACCGTCCACCACTGCTTTGCCACCAACCGAGACTGCGC 479
   |||||
```

Db 171735 TCGAGCTGAGCTTTGCTGCGCCATCCACACAGACTGTTTGCGCCCTGCGACGCCGCTG 171794

Qy 480 CTGACTCCCATCCCTCTGGAT 500

Db 171795 CTGACTCCCATCCCTCTCGAGAT 171815

Search completed: April 17, 2003, 06:37:59
Job time : 2080.07 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 02:30:00 ; Search time 180.426 Seconds
(without alignments)
6240.778 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500
Sequence: 1 cccctgggcggcgtccctcctt.....tgaccctccatccctctgcat 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	500	100.0	10499	21	ABN97929 Human retroviral s
2	500	100.0	56093	24	ABL61744 Colon adenocarcino
3	435	87.0	711	21	ABN97946 Human retroviral s
4	413.4	82.7	7466	23	AA568626 DNA encoding novel
5	410	82.0	783	20	AA525669 Human endogenous r
6	383.8	76.8	711	21	ABN97947 Human retroviral s
7	348.2	69.6	1329	19	AAV43219 Multiple sclerosis
8	348.2	69.6	1329	20	AA529704 Clone 5M6 from MSR
9	326.8	65.4	1393	22	AA531002 Human diagnostic a

10	306.8	61.4	893	23	AA565964	DNA encoding novel
11	296.2	59.2	849	22	AA531000	Human diagnostic a
12	295.6	59.1	1478	23	AA587568	DNA encoding novel
13	292.4	58.5	808	23	AA588392	DNA encoding novel
14	290.8	58.2	1243	23	AA584189	DNA encoding novel
15	287.6	57.5	2629	22	AAK87491	Human immune/haema
16	286.4	57.3	2781	22	AA555630	Nucleotide sequenc
17	284.8	57.0	2946	22	AA577526	Human secreted pro
18	284.6	56.9	2300	23	AA525468	Human secreted pro
19	284.6	56.9	2300	23	AA592680	DNA encoding novel
20	283.8	56.8	2830	24	AA524195	Human syncytin cdn
21	283.2	56.6	3372	21	AA525663	Human endogenous r
22	283.2	56.6	3372	21	AA559213	Partial pol gene a
23	283.2	56.6	7582	20	AA525665	Complete human end
24	283.2	56.6	7582	21	AA559215	Human endogenous r
25	282.2	56.4	3831	23	AA571727	DNA encoding novel
26	282.2	56.4	5154	23	AA567609	DNA encoding novel
27	282.2	56.4	8279	23	AA576474	DNA encoding novel
28	282.2	56.4	8294	23	AA584209	DNA encoding novel
29	281.6	56.3	2782	20	AA525661	Human endogenous r
30	281.6	56.3	2782	21	AA559211	5' non coding, 3'
31	281.6	56.3	2782	22	AA520069	HERV-W envelope, pr
32	280.8	56.2	1165	23	AA572721	DNA encoding novel
33	280.6	56.1	1136	20	AA525660	Human endogenous r
34	280.6	56.1	1136	21	AA559210	3' pol gene and 3'
35	280	56.0	2942	22	AA577313	DNA encoding novel
36	279.2	55.8	17758	22	AA584138	Human immune/haema
37	277.6	55.5	416	23	AA565963	DNA encoding novel
38	270.4	54.1	635	19	AAV43215	Multiple sclerosis
39	270.4	54.1	635	20	AA529702	Clone Cl6-3' from
40	270.4	54.1	2030	21	AA563826	Nucleotide sequenc
41	261.4	52.3	410	24	ABN94040	Gene #538 used to
42	261.4	52.3	410	24	ABK64210	Human benign prosc
43	259.6	51.9	448	23	AA587567	DNA encoding novel
44	259.6	51.9	583	23	AA588391	DNA encoding novel
45	259.2	51.8	541	22	AA562368	Human foetal liver

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
XX ABN97929;
AC
XX
XX 01-AUG-2002 (first entry)
DT
XX
XX Human retroviral sequence HERV-7q.
DE
XX
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX
XX Human retrovirus.
OS
XX
XX WO967395-A1.
PN
XX
XX 29-DEC-1999.
PD
XX
XX 23-JUN-1999; 99WO-FR01513.
PF
XX
XX 23-JUN-1998; 98FR-0007920.
PR
XX
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA
XX
XX Alliel PM, Perin J, Rieger F;
PI
XX
XX WPI; 2000-160587/14.
DR
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225bp; French.
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
XX
Query Match 100.0%; Score 500; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 5; le-164;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 CCTTGGGGGCGGGCTTCTTCTTGGGATGAGGCAAAAGCTGAGATACAGCAATTATC 60
DB 1 CCTGGGGCGGGCTTCTTCTTGGGATGAGGCAAAAGCTGAGATACAGCAATTATC 60
QY 61 TTGGAACGTGAG 120
DB 61 TTGGAACGTGAG 120
QY 121 GCTGGGAAGGTGACACACCTCCACTTTAAACACGGGGCTTGCAACTTATGCTACAGCTGA 180
DB 121 GCTGGGAAGGTGACACACCTCCACTTTAAACACGGGGCTTGCAACTTATGCTACAGCTGA 180
QY 181 CCAATCAGAGAGCTCCTAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCA 240
DB 181 CCAATCAGAGAGCTCCTAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCA 240
QY 241 ATCATTCATTCGCTGAG 300
DB 241 ATCATTCATTCGCTGAG 300
QY 301 TCGAGCTGGCAACAGACAG 360
DB 301 TCGAGCTGGCAACAGACAG 360
QY 361 TGTATTTCACTATTAATTTGCAACTGCACTCTTTGGTCCATGTTCTTACGAGCT 420
DB 361 TGTATTTCACTATTAATTTGCAACTGCACTCTTTGGTCCATGTTCTTACGAGCT 420
QY 421 CGAGCTGAGCTTTTGGCTGACGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 CGAGCTGAGCTTTTGGCTGACGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 TGACTCCCATCCCTCTGGAT 500
DB 481 TGACTCCCATCCCTCTGGAT 500
XX
RESULT 2
ABL61744
ID ABL61744 standard; DNA; 56093 BP.
XX
AC ABL61744;
XX
XX 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM Stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytosstatic; gene therapy; anti-neoplastic; Wilms' tumour; adenocarcinoma;
CC

KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 22-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
XX (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppe DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 81; 44bp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytosstatic

CC activity and can be used in gene therapy. MI can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of MI, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. MI can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophagial, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 56093 BP, 16164 A, 12346 C, 10702 G, 16881 T, 0 other;

Query Match 100.0%; Score 500; DB 24; Length 56093;

Best Local Similarity 100.0%; Pred. No. 1.3e-163;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTGGGGCGGGCTCTCTTCTTGAGTACGGCAAAACGCTGAGATACAGCAATTATC 60
 DB 28001 CCTTGGGGCGGGCTCTCTTCTTGAGTACGGCAAAACGCTGAGATACAGCAATTATC 28060
 QY 61 TTGCACTGAGAGACAGGACTAGTGGATTCTTCTAGGCGGCACTAAGATCCCTAAGCCTA 120
 DB 28061 TTGCACTGAGAGACAGGACTAGTGGATTCTTCTAGGCGGCACTAAGATCCCTAAGCCTA 28120
 QY 121 GCTGGGAAGGTGACCAAGTCCACCTTTTAAACAGCGGGCTTGCAACTTATGCTACACCTGA 180
 DB 28121 GCTGGGAAGGTGACCAAGTCCACCTTTTAAACAGCGGGCTTGCAACTTATGCTACACCTGA 28180
 QY 181 CCAATCAGAGAGCTCATAAATGCTAATTAGCCAAAGACAGAGTAAAGAAATAGCCA 240
 DB 28181 CCAATCAGAGAGCTCATAAATGCTAATTAGCCAAAGACAGAGTAAAGAAATAGCCA 28240
 QY 241 ATCATATTGCTCGAGAGCAGAGAGGAGCAACAATGGGATATAAACCAGGAT 300
 DB 28241 ATCATATTGCTCGAGAGCAGAGAGGAGCAACAATGGGATATAAACCAGGAT 28300
 QY 301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCCTTGTATGGAGACTGTTTTCA 360
 DB 28301 TCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCCTTGTATGGAGACTGTTTTCA 28360
 QY 361 TGTATTATCTATTAATATCTTGCAACTGCACTCTTGTGCTCCATGTTTCTTAAGGCT 420
 DB 28361 TGTATTATCTATTAATATCTTGCAACTGCACTCTTGTGCTCCATGTTTCTTAAGGCT 28420
 QY 421 CGAGCTGAGCTTTTGTCTACCGTCCACACTGCTGTTTGGCAACACCGCAACCTGCGGC 480
 DB 28421 CGAGCTGAGCTTTTGTCTACCGTCCACACTGCTGTTTGGCAACACCGCAACCTGCGGC 28480
 QY 481 TGACTCCCATCCCTCTGGAT 500
 DB 28481 TGACTCCCATCCCTCTGGAT 28500

RESULT 3

ID AEN97946 standard; DNA; 711 BP.

XX AEN97946;

XX 01-AUG-2002 (first entry)

XX Human retroviral sequence RL.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX Multiple sclerosis; ds.

XX Human retrovirus.

XX WO967395-A1.

XX 29-DEC-1999.

PF 23-JUN-1999; 99WO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,

PT used for diagnosis, treatment and prevention of autoimmune and

PS neurological diseases -

CC Claim 3; Fig 3; 225pp; French.

CC The present invention relates to new nucleic acid sequences of human

CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.

CC Regulatory elements associated with HERV-7q may alter expression of other

CC genes (even remote genes) on the same chromosome, inducing immunological

CC and/or neurological changes (which may be pathological or protective/

CC curative). HERV-7q peptides can be used to improve efficiency of the

CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding

CC sequences can be used in immunogenic or vaccinating compositions, for

CC protection against autoimmune diseases, particularly multiple sclerosis.

CC The peptides may also be used (by sequence comparison) to detect/identify

CC endogenous retroviruses that are abnormally expressed in cancer,

CC neuropathologies or other autoimmune diseases. The present sequence was

XX used to illustrate the invention.

XX Sequence 711 BP, 170 A, 204 C, 162 G, 175 T, 0 other;

Query Match 87.0%; Score 435; DB 21; Length 711;

Best Local Similarity 100.0%; Pred. No. 7.1e-142;

Matches 435; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 ACTGAGAGACAGACTAGTGGATTCTTCTAGCGCGCACTAAGATCCCTAAGCTAGCTG 125
 DB 1 ACTGAGAGACAGACTAGTGGATTCTTCTAGCGCGCACTAAGATCCCTAAGCTAGCTG 60
 QY 126 GAAGGTGACCAAGTCCACCTTTTAAACAGGGGCTTGCAACTTATGCTACACCTTACCAAT 185
 DB 61 GAAGGTGACCAAGTCCACCTTTTAAACAGGGGCTTGCAACTTATGCTACACCTTACCAAT 120
 QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTTAAAGAAATAGCAATCAT 245
 DB 121 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTTAAAGAAATAGCAATCAT 180
 QY 246 CTATTGCTGAGACAGACAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 305
 DB 181 CTATTGCTGAGACAGACAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 240
 QY 306 CTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCCTTGTATGGAGCTGTTTCATGCTA 365
 DB 241 CTGGCAACAGACAGCCCCCTTTGGGTCCCTTCCCTTGTATGGAGCTGTTTCATGCTA 300
 QY 366 TTTCACTCTAATTAATCTTGCAACTGCACTCTTGTGCAATGTTTCTTAAGGCTCGAGC 425
 DB 301 TTTCACTCTAATTAATCTTGCAACTGCACTCTTGTGCAATGTTTCTTAAGGCTCGAGC 360
 QY 426 TGAGCTTTTGTCTACCGTCCACACTGCTGTTTGGCAACACCGGACAGACTGCTGACT 485
 DB 361 TGAGCTTTTGTCTACCGTCCACACTGCTGTTTGGCAACACCGGACAGACTGCTGACT 420
 QY 486 CCGATCCCTCTGGAT 500
 DB 421 CCGATCCCTCTGGAT 435

RESULT 4

XX AAS68626 standard; cDNA; 7466 BP.

XX

AA68626;
13-FEB-2002 (first entry)
DNA encoding novel human diagnostic procain #4430.
Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
Homo sapiens.
MO200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Dzmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
P-PSDB; ABG04439.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 1; SEQ ID No 4430; 103bp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;
Query Match 82.7%; Score 413.4; DB 23; Length 7466;
Best Local Similarity 98.2%; Pred. No. 1e-133;
Matches 429; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

66 ACTGAGAGACAGACTAGTGGATTTCTTAGGCGGACTAGATCCCTAGGCTAGCTGG 125
3716 ACTGAGAGACAGACTAGTGGATTTCTTAGGCGGACTAGATCCCTAGGCTAGCTGG 3775
126 GAAGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
3776 GAAGTGACACGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 3835
186 CAGAGAGCTCACTAAATGCTATTAGGCAAGACAGGGGGTAAAGAAATAGCAATCAT 245
3836 CAGAGAGCTCACTAAATGCTATTAGGCAAGACAGGGGGTAAAGAAATAGCAATCAT 3895

246 CTATTGCTGAGACACAGCAGAGGGAACAATCGGATATTAACCCAGGCAATTCGAG 305
3896 TTATTGCTGAGACACAGCAGAGGGAACAATCGGATATTAACCCAGGCAATTCGAG 3955
306 CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCCTTTGTATAGGAG--CTGTTTCAATGC 363
3956 CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCCTTTGTATAGGAGCTGTTTCAATGC 4015
364 TATTCACTATTAAATCTTGAACAGTCACTCTTGTGTCATGTTCTTACGAGCTGA 423
4016 TATTCACTATTAAATCTTGAACAGTCACTCTTGTGTCATGTTCTTACGAGCTGA 4075
424 GCTGAGCTTTTGTCTACCGGTCACCACTGCTTTTGCACACCGCAGACCTGCCGCTGA 483
4076 GCTGAGCTTTTGTCTACCGGTCACCACTGCTTTTGCACACCGCAGACCTGCCGCTGA 4135
484 CTCCCATCCCTCTGGAT 500
4136 CTCCCATCCCTCTGGAT 4152

RESULT 5
AAX25669
ID AAX25669 standard; cDNA to mRNA; 783 BP.
AAX25669;
21-MAY-1999 (first entry)
Human endogenous retrovirus W long terminal repeat region.
Clone; human endogenous retrovirus; genome; autoimmune disease;
multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
Human endogenous retrovirus.
WO9902696-A1.
21-JAN-1999.
06-JUL-1998; 98WO-FR01442.
07-JUL-1997; 97FR-0008815.
(INMR) BIO MERIEUX.
Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
WPI; 1999-120897/10.
New nucleic acid sequences from human endogenous retrovirus-W -
expressed exclusively in placenta and useful in diagnosis and
therapy of autoimmune disease, and abnormal or failed pregnancy
Claim 1; Page 83; 106bp; French.
This sequence represents the long terminal repeat (LTR) region of the
human endogenous retrovirus (HERV) W genome. The nucleic acids, their
fragments or peptides encoded by them are markers of autoimmune disease
(e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus
erythematosus, insulin-dependent diabetes and related pathologies) and
of abnormal or unsuccessful pregnancy and can be used as chromosomal
markers for susceptibility to these conditions, or proximity markers
of genes associated with this susceptibility.
SQ Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;
Query Match 82.0%; Score 410; DB 20; Length 783;
Best Local Similarity 93.8%; Pred. No. 4.5e-133;
Matches 408; Conservative 24; Mismatches 1; Indels 2; Gaps 1;

QY 68 TGAAGACAGAGCTAGTGGATTTCCTAGGCCGACATAAGAAATCCCTAAGCTAGCTGGGA 127
DB 1 TGAAGACAGAGCTAGTGGATTTCCTAGGCCGACATAAGAAATCCCTAAGCTAGCTGGGA 60
QY 128 AGGAGACAGCTCCAGCTTTAAACAGGGGGCTTGAATTTAGCTCAGACCTGACCAATCA 187
DB 61 AGGAGACAGCTCCAGCTTTAAACAGGGGGCTTGAATTTAGCTCAGACCTGACCAATCA 120
QY 188 GAGAGCTCACTAAATGCTAATTTAGCAAAAGACAGAGTAAAGAAATAGCCATCATCT 247
DB 121 GAGAGCTCACTAAATGCTAATTTAGCAAAAGACAGAGTAAAGAAATAGCCATCATCT 180
QY 248 ATTGCTTGAAGACAGACAGAGGAGCAACAATCGGATATTAACCCAGGATTCGAGCT 307
DB 181 ATTGCTTGAAGACAGACAGAGGAGCAACAATCGGATATTAACCCAGGATTCGAGCT 240
QY 308 GGAACAGACAGACCCCTTTGGGTCCTTCCCTTTGATGGAG--CTGTTTCATGCTA 365
DB 241 GGAACAGACAGACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCATGCTA 300
QY 366 TTTCACTCTAATTAATCTTGAACCTGACCTCTTGTGTCATGTTTCTTACGGCTGAGC 425
DB 301 TTTCACTCTAATTAATCTTGAACCTGACCTCTTGTGTCATGTTTCTTACGGCTGAGC 360
QY 426 TGAAGCTTTGCTCAGCGTCCACCACTGCTTTTGCAACACCGAGACTGCGCTGACT 485
DB 361 TGAAGCTTTGCTCAGCGTCCACCACTGCTTTTGCAACACCGAGACTGCGCTGACT 420
QY 486 CCCATCCCTCTGGAT 500
DB 421 CCCATCCCTCTGGAT 435

RESULT 6

ABN97947

ID ABN97947 standard; DNA; 711 BP.

AC ABN97947;

DT 01-AUG-2002 (first entry)

DE Human retroviral sequence RIF.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

KM multiple sclerosis; ds.

XX Human retrovirus.

XX MO9967395-A1.

XX 29-DEC-1999.

XX 23-JUN-1999; 99MO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,

PT used for diagnosis, treatment and prevention of autoimmune and

XX neurological diseases -

XX Claim 3; Fig 3; 225BP; French.

CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/

CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.

SQ Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;

Query Match 76.8%; Score 383.8; DB 21; Length 711;

Best Local Similarity 95.4%; Pred. No. 6.6e-124;

Matches 417; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGCACTAGCTGATTTCTAGGCCGACATAAGAAATCCCTAAGCTAGCTGG 125
DB 1 ACTGAGAGACAGCACTAGCTGATTTCTAGGCCGACATAAGAAATCCCTAAGCTAGCTGG 60
QY 126 GAAGGTACACACGCTCCAGCTTTAAACAGGGGGCTTGAACCTTAGCTCAGACCTGACCAAT 185
DB 61 GAAGGTACACACGCTCCAGCTTTAAACAGGGGGCTTGAACCTTAGCTCAGACCTGACCAAT 120
QY 186 CAGAGAGCTCACTAAATGCTAATTTAGCAAAAGACAGAGTAAAGAAATAGCCATCAT 245
DB 121 CAGAGAGCTCACTAAATGCTAATTTAGCAAAAGACAGAGTAAAGAAATAGCCATCAT 180
QY 246 CTATTGCTGAGAGACAGAGGAGGAGCAACAATCGGATATTAACCCAGGACTTTGGAG 305
DB 181 CTATTGCTGAGAGACAGAGGAGGAGCAACAATCGGATATTAACCCAGGACTTTGGAG 240
QY 306 CTGCAACAGAGAGCCCTTTGGGTCCTTCCCTTTGATGGAG--CTGTTTCATGAC 363
DB 241 CTGCAACAGAGAGCCCTTTGGGTCCTTCCCTTTGATGGAGCTGTTTCATGAC 299
QY 364 TATTCACTCTAATTAATCTTGAACCTGACCTTGTGTCATGTTTCTTACGGCTGGA 423
DB 300 TATTCACTCTAATTAATCTTGAACCTGACCTTGTGTCATGTTTCTTACGGCTGGA 359
QY 424 GCTGAGCTTTGCTCAGCGTCCACCACTGCTTTTGCAACACCGAGACTGCGGCTGGA 483
DB 360 GCTGAGCTTTGCTCAGCGTCCACCACTGCTTTTGCAACACCGAGACTGCGGCTGGA 419
QY 484 CTCCCATCCCTCTGGAT 500
DB 420 CTCCCATCCCTCTGGAT 436

RESULT 7

AAV43219

ID AAV43219 standard; cDNA; 1329 BP.

AC AAV43219;

DT 29-DEC-1998 (first entry)

DE Multiple sclerosis associated retrovirus fragment 6.

XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;

XX gag gene; env gene; rheumatoid arthritis-associated virus; ss.

XX Multiple sclerosis associated retrovirus.

XX Key Location/Qualifiers

XX CDS 2..490

XX /*tag= a

XX /product= "Encodes protein AAV71069"

XX /transl_except= (pos:77-79, appears to code for a
XX stop codon)
XX /transl_except= (pos:125-127, appears to code for a
XX stop codon)
XX /transl_except= (pos:137-139, appears to code for a
XX stop codon)

XX MO9823755-A1.
XX 04-JUN-1998.
XX 26-NOV-1997; 97WO-IB01482.
XX 26-NOV-1996; 96US-0756429.
XX (INMR) BIO MERIEUX.
XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F,
XX Mandrand B, Paranhos-Baccala G, Perron H;
XX WPI; 1998-322732/28.
XX P-PSDB; AAW71069.
XX New nucleic acid from retroviruses - useful for diagnosis,
XX prevention and treatment of, e.g. multiple sclerosis
XX
XX Disclosure; Pages 187-188; 286pp; English.
XX
XX The present sequence represents a multiple sclerosis (MS) associated
XX retrovirus (MSRV) genomic fragment used in the method of the
XX invention. The invention provides complete or partial genomic
XX sequences of the MSRV-1 pol gene, gag gene, and env gene, and
XX polypeptides encoded by these genes. The invention also provides
XX antibodies raised against the polypeptides. The genomic sequences,
XX polypeptides and antibodies are also claimed useful for diagnosing,
XX infection by MS and rheumatoid arthritis-associated viruses, and also
XX for prevention and treatment of infection with these viruses.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 69.6%; Score 348.2; DB 19; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCGGACPAAGATCCCTAAGCTTAGCTGG 125
DB 547 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCGGACPAAGATCCCTAAGCTTAGCTGG 606
QY 126 GAAGGTGACACGCTCCACCTTTAAACACGGGGCTTGCACCTTAGCTCACACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCACCTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTTAATAGGCAAGACAGAGTAAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTTAATAGGCAAGACAGAGTAAAGCAATTAAGCCATCAT 726
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACATCGCATTAATTAACCCAGGCAATTGAG 305
DB 727 CTATTGCTGAGAGACAGACAGGAGGAGCAAGAGATTGGATTAATTAAGGCAATTCAAG 786
QY 306 CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCCTTTGATGGAG--CTGTTTTCAAGC 363
DB 787 CCAGAGACAGCA-ACCCCCCTTTGGGTCCCTTCCCATTTGATGGAGCTGTGTTTCACTC 845
QY 364 TATTTCACCTTAATTAATTTTGAACCTGCACTCTTGTGTCCATGTTTCTTAACGGCTGA 423
DB 846 TATTTCACCTTAATTAATTAATGCAACCTGCACTCTTGTGTCCATGTTTCTTAAGCTCA 905
QY 424 GGTGAGCTTTTGTCTACCGTCCACCACTGCTGTTGGCCACACCGGAGACCTGCGCTGA 483
DB 906 GGTGAGCTTTTGTGTCCTCCACCACTGCTGTTGGCCACCGTCCAGACCGGCTGCTGA 965
QY 484 CTCCCATCCCTCTGGAT 500
DB 966 CTTCATCCCTTTGGAT 982

RESULT 8
AA29704

ID AA29704 standard; DNA; 1329 BP.
XX AAX29704;
XX 08-JUN-1999 (first entry)
XX Clone 5M6 from MSRV-1.
XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
XX rheumatoid polyarthritis; ss.
XX
XX Multiple sclerosis related virus type 1.
XX
XX FR2765588-A1.
XX 08-JAN-1999.
XX
XX 07-JUL-1997; 97FR-0008816.
XX 07-JUL-1997; 97FR-0008816.
XX 07-JUL-1997; 97FR-0008816.
XX (INMR) BIO MERIEUX.
XX WPI; 1999-098275/09.
XX P-PSDB; AAW99554.
XX
XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
XX multiple sclerosis or rheumatoid polyarthritis
XX
XX Claim 1; Page 39-40; 83pp; French.
XX
XX This sequence represents clone 5M6 from a novel multiple sclerosis
XX related virus type 1 (MSRV1). The sequence can be used in diagnostic,
XX prophylactic or therapeutic compositions to inhibit expression of a
XX multiple sclerosis related virus and/or virus associated with
XX rheumatoid polyarthritis.
XX
XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 69.6%; Score 348.2; DB 20; Length 1329;
Best Local Similarity 90.2%; Pred. No. 3e-111;
Matches 394; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCGGACPAAGATCCCTAAGCTTAGCTGG 125
DB 547 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCGGACPAAGATCCCTAAGCTTAGCTGG 606
QY 126 GAAGGTGACACGCTCCACCTTTAAACACGGGGCTTGCACCTTAGCTCACACCTGACCAAT 185
DB 607 GAAGGTGACCGCATCCATCTTTAAACATGGGGCTTGCACCTTAGCTCACACCGACCAAT 666
QY 186 CAGAGAGCTCACTAAATGCTTAATAGGCAAGACAGAGGTAAGAAATAGCCATCAT 245
DB 667 CAGAGAGCTCACTAAATGCTTAATAGGCAAGAAACAGAGTAAAGCAATTAAGCCATCAT 726
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACATCGGATTAATTAACCCAGGCAATTGAG 305
DB 727 CTATTGCTGAGAGACAGCGGAGGAGCAAGAGATTGGATTAATTAAGGCAATTCAAG 786
QY 306 CTGGCAACAGACGCCCCCTTTGGGTCCCTTCCCTTTGATGGAG--CTGTTTTCAAGC 363
DB 787 CCAGAGACAGCA-ACCCCCCTTTGGGTCCCTTCCCATTTGATGGAGCTGTGTTTCACTC 845
QY 364 TATTTCACCTTAATTAATTTTGAACCTGCACTCTTGTGTCCATGTTTCTTAACGGCTGA 423
DB 846 TATTTCACCTTAATTAATTAATGCAACCTGCACTCTTGTGTCCATGTTTCTTAAGCTCA 905
QY 424 GGTGAGCTTTTGTCTACCGTCCACCACTGCTGTTGGCCACACCGGAGACCTGCGCTGA 483
DB 906 GGTGAGCTTTTGTGTCCTCCACCACTGCTGTTGGCCACCGTCCAGACCGGCTGCTGA 965
QY 484 CTCCCATCCCTCTGGAT 500

Db 966 CTTCCATCCTTGAT 982

RESULT 9

AAS31002
ID AAS31002 standard; cDNA; 1393 BP.

XX AAS31002;

XX 04-DEC-2001 (first entry)

DE Human diagnostic and therapeutic polynucleotide (DITHP) #17.

XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;

KM cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;

KM acquired immune deficiency syndrome; AIDS; autoimmune disorder;

XX respiratory disorder; ss.

XX Homo sapiens.

XX WO200162927-A2.

XX 30-AUG-2001.

XX 21-FEB-2001; 2001WO-US06059.

XX 24-FEB-2000; 2000US-0184693.

XX 24-FEB-2000; 2000US-0184697.

XX 24-FEB-2000; 2000US-0184768.

XX 24-FEB-2000; 2000US-0184769.

XX 24-FEB-2000; 2000US-0184770.

XX 24-FEB-2000; 2000US-0184771.

XX 24-FEB-2000; 2000US-0184772.

XX 24-FEB-2000; 2000US-0184773.

XX 24-FEB-2000; 2000US-0184774.

XX 24-FEB-2000; 2000US-0184776.

XX 24-FEB-2000; 2000US-0184777.

XX 24-FEB-2000; 2000US-0184797.

XX 24-FEB-2000; 2000US-0184813.

XX 24-FEB-2000; 2000US-0184837.

XX 24-FEB-2000; 2000US-0184841.

XX 24-FEB-2000; 2000US-0185213.

XX 12-MAY-2000; 2000US-0203785.

XX 15-MAY-2000; 2000US-0204226.

XX 16-MAY-2000; 2000US-0204525.

XX 16-MAY-2000; 2000US-0204821.

XX 16-MAY-2000; 2000US-0204908.

XX 16-MAY-2000; 2000US-0205332.

XX 17-MAY-2000; 2000US-0204815.

XX 17-MAY-2000; 2000US-0204863.

XX 17-MAY-2000; 2000US-0205221.

XX 17-MAY-2000; 2000US-0205285.

XX 17-MAY-2000; 2000US-0205286.

XX 17-MAY-2000; 2000US-0205287.

XX 17-MAY-2000; 2000US-0205323.

XX 17-MAY-2000; 2000US-0205324.

XX (INCY-) INCYTE GENOMICS INC.

XX Panzer A, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;

XX Chen A, D'Sa SA, Amehy S, Dahl CR, Dam TC, Daniels SE;

XX Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;

XX Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockbrecher TK, Dafo A;

XX Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;

XX Cohen HJ, Hodgson DM, Lincoln SE, Jackson SJ;

XX MPI; 2001-502867/55.

XX P-PSDB; AAU19431.

XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.

XX enzymes, hormones and receptors, useful in diagnostics and therapeutics

PT -
XX Claim 1; Page 304; 522pp; English.

CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC assay (ELISA)). AAS30986-AAS31196 represent human diagnostic and
CC therapeutic (DITHP) polynucleotides of the invention.

SQ Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 2 other;

Query Match 65.4%; Score 326.8; DB 22; Length 1393;
Best Local Similarity 88.7%; Pred. No. 1e-103;
Matches 394; Conservative 0; Mismatches 32; Indels 18; Gaps 3;

QY 66 ACTGAGAGACAGACTGATGATTTCTAGGCGGACCTAAGATCCCTAGCTAGCTGG 125

Db 245 ACTGAGAGAGAGACTGATGATTTCTAGGCGGACCTAAGATCCCTAGCTAGCTGG 304

QY 126 GAAAGTGACCACTGCTTAAACAGGGGCTTGAACCTTAGCTACACCTGACCAAT 185

Db 305 GAAAGTGACCGCGCTTAAACAGGGGCTTGAACCTTAGCTACACCTGACCAAT 364

QY 186 C-----AGAGACTGACTAATAATGCTAATTAAGCAAAAGCAGAGGTAAGAAATA 236

Db 365 CAGGTAGTAAGAGAGCTCACTAAATGCTAATTAAGCAAAAGCAGAGGTAAGAAATA 424

QY 237 GCCAATCATCTATTGCTGAGACACAGACAGAGGACCAATCGGATATTAACCCAG 296

Db 425 GCCAATCATCTATTGCTGAGACACAGACAGAGGACCAATCGGATATTAACCCAG 484

QY 297 GCATTGAGCTGGCAACAGACGCCCTTTGGGTCCTTCCCTTTGATAGGAGCTGTT 356

Db 485 GCATTGAGCTGGCAACAGACGCCCTTTGGGTCCTTCCCTTTGATAGGAGCTGTT 543

QY 357 TTCAATGCTATTACCTCTTAAATCTTGCAACTGACACTTCTGATCATGTTCTTAC 416

Db 544 T-----TTTCACTCTATTAAATCTTGCAACTGACACTTCTGATCATGTTCTTAC 595

QY 417 GGCTGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTGTTTCCACCCAGAGACTG 476

Db 596 GGTTTGAAGCTGAGCTTTTGTCTACCGTCCACCACTGCTGTTTCCACCCAGAGACTG 655

QY 477 CGGCTGACTCCCATCCCTCTTGAT 500

Db 656 CGGCTGACTCCCATCCCTCTTGAT 679

RESULT 10

AAS65964
ID AAS65964 standard; cDNA; 893 BP.

XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.

XX enzymes, hormones and receptors, useful in diagnostics and therapeutics

AA65964;
13-FEB-2002 (first entry)
DNA encoding novel human diagnostic protein #1768.
Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder; ss.
Homo sapiens.
W0200175067-A2.
11-OCT-2001.
30-MAR-2001; 2001WO-US08631.
31-MAR-2000; 2000US-0540217.
23-AUG-2000; 2000US-0649167.
(HYSE-) HYSEQ INC.
Dmanac RT, Liu C, Tang YT;
WPI; 2001-639362/73.
P-PSDB; ABG01777.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 1; SEQ ID No 1768; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridization probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pat_sequences.
Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;
Query Match 61.4%; Score 306.8; DB 23; Length 893;
Best Local Similarity 87.2%; Pred. No. 8.1e-97;
Matches 389; Conservative 0; Mismatches 37; Indels 20; Gaps 4;
66 ACTGAGAGCAGAGCTAGTGGATTCTTAGGCCGACTAGAAATCCCTAAGCTAGCTGG 125
DB 236 ACTGAGAGCAGAGCTAGTGGATTCTTAGGCCGACTAGAAATCCCTAAGCTAGCTGG 295
126 GAAGGTGACACGCTCCTTTAAACACGGGCTTGCA/CTTAGCTCACACCTGACCAAT 185
DB 236 GAAGGTGACACGCTCCTTTAAACACGGGCTTGCA/CTTAGCTCACACCTGACCAAT 355
186 C-----AAGAGCTACTAAATGCTAATTAGGCTAAGACAGAGAGTAAAGAAATA 236
DB 356 CAGGTAAAGAAAGAGCCGCTAAATGCTAATTAGGCTAAGACAGAGTAAAGAAATA 415

237 GCCAATCATCTATTGCTGCTGAGACACAGAGGAGACAAATCGGATATAAACCAG 296
DB 416 GCCAATCATCTATTGCTGCTGAGACACAGAGGAGGAGCAATGATCAATATAACCCAG 475
297 GCATTGAGCTGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGTAATGAGAGCTGTT 356
DB 476 GCATTGAGCTGCAACAGACAGCCCTTTGGGTCCTTCCCTTTGTAATGAGAGCTGTT 534
357 TTGATGCTATTTCAGTCTATTAATCTTGCACTG--CACTCTTGCTGCTCAATGTTCTT 414
DB 535 T-----TTTCACTATTAATCTTGCACTGCAACTCTTGCTGCTGTTGTTGTT 586
415 ACGGCTGAGCTGAGCTTTTGTCTCACCGTCCAGCACTGCTGTTGCCACCGAGACC 474
DB 587 ACGGCTGAGCTGAGCTTTTGTCTCACCGTCCAGCACTGCTGTTGCCACCGAGACC 646
475 TGGCGTGAAGCTCCCATCCCTCTGGAT 500
DB 647 CATGCTGAAGCTCCCATCCCTCTGGAT 672
RESULT 11
AAS31000
ID AAS31000 standard; cDNA; 849 BP.
AC AAS31000;
XX AAS31000;
DT 04-DEC-2001 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide (DITHP) #15.
XX
XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
XX cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
XX acquired immune deficiency syndrome; AIDS; autoimmune disorder;
XX respiratory disorder; ss.
OS
XX Homo sapiens.
XX
XX W0200162927-A2.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-US06059.
XX
XX 24-FEB-2000; 2000US-0184693.
XX 24-FEB-2000; 2000US-0184697.
XX 24-FEB-2000; 2000US-0184698.
XX 24-FEB-2000; 2000US-0184768.
XX 24-FEB-2000; 2000US-0184769.
XX 24-FEB-2000; 2000US-0184770.
XX 24-FEB-2000; 2000US-0184771.
XX 24-FEB-2000; 2000US-0184772.
XX 24-FEB-2000; 2000US-0184773.
XX 24-FEB-2000; 2000US-0184774.
XX 24-FEB-2000; 2000US-0184776.
XX 24-FEB-2000; 2000US-0184777.
XX 24-FEB-2000; 2000US-0184779.
XX 24-FEB-2000; 2000US-0184813.
XX 24-FEB-2000; 2000US-0184837.
XX 24-FEB-2000; 2000US-0184841.
XX 24-FEB-2000; 2000US-0185213.
XX 24-FEB-2000; 2000US-0185216.
XX 12-MAY-2000; 2000US-0203785.
XX 15-MAY-2000; 2000US-0204226.
XX 16-MAY-2000; 2000US-0204525.
XX 16-MAY-2000; 2000US-0204821.
XX 16-MAY-2000; 2000US-0204908.
XX 16-MAY-2000; 2000US-0205232.
XX 17-MAY-2000; 2000US-0204815.
XX 17-MAY-2000; 2000US-0204863.
XX 17-MAY-2000; 2000US-0205221.
XX 17-MAY-2000; 2000US-0205285.

CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1478 BP; 389 A; 337 C; 364 G; 380 T; 0 other;
XX
Query Match 59.1%; Score 295.6; DB 23; Length 1478;
Best Local Similarity 85.7%; Pred. No. 9,76-93;
Matches 382; Conservative 0; Mismatches 44; Indels 20; Gaps 4;
XX
66 ACTGAGAGACGAGCTAGCTGATTTCTTAGGCGCACTTAAGATCCCTAAGCTTAGCTG 125
DB 1473 ATTAAAGACGAGACGACCTGATGTCTTAGGCTGATTAAGATCCCTAAGCTTAGCTG 1414
XX 126 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGCACCTAGCTCACTTACCAAT 185
DB 1413 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGCACCTAGCTCACTTACCAAT 1354
XX 186 C-----AGAGAGCTCACTTAAGCTTAAGTGAAGACAGAGAGGTTAAAGAAATA 236
DB 1353 CAGGTAGTAAAGAGAGCTCACTTAAGGCTTAAGGCAAAAGAGAGGTTAAAGAAATA 1294
XX 237 GCCAATCATCTATTGCTGAGAGACAGACAGAGAGGACCAACATCGGGATTAAGCCAG 296
DB 1293 GCCAATCATCTATTGCTGAGAGACAGACAGAGAGGACCAACATCGGGATTAAGCCAG 1234
XX 297 GCATTGAGCTGGCAAGACAGACGCCCCCTTGGGCTCCCTCCCTTTGATGAGAGCTGTT 356
DB 1233 GCATTGAGCTGGCAAGAGGC-TACCCCTCTGAGGCTCACTCCCTTTGATGAGAGCTCTG 1175
XX 357 TTCAATGCTATTCTCACTTAATTAATCTTGCACATG--TACTCTTGTGCTCATGTTCTT 414
DB 1174 T-----TTTCACTCTATTAAATCTTGCACATGCACTCTTGTGCTCATGTTCTT 1123
XX 415 ACGGCTGAGCTGAGCTTTTGTGCTACCGTCCACGACTCTGTGTTGCCACCGAGACC 474
DB 1122 ATGGCTCAGCTGAGCTTTGCTGCTGACATCCACGACTCTGTGTTGCCACCGAGACC 1063
XX 475 TGCCGCTGACTCCCATCCCTCTGAGT 500
DB 1062 CGTAACTGACTTCCACCCCTCCGAGT 1037
XX
RESULT 13
AAS88392
ID AAS88392 standard; cDNA; 808 BP.
XX
AAS88392;
AC
XX 13-FEB-2002 (first entry)
DT
XX
DNA encoding novel human diagnostic protein #24196.
DE
XX
Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM Food Supplement; medical imaging; diagnostic; genetic disorder; ss.
KM
XX
Homo sapiens.
OS
XX
MO200175067-A2.
PN
XX
11-OCT-2001.
PD
XX
30-MAR-2001; 2001WO-US08631.
PF
XX
31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PA
XX (HYSE-) HYSEQ INC.
PI
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
XX P-PSDB; A8624205.
PT
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 24196; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 808 BP; 184 A; 234 C; 189 G; 201 T; 0 other;
XX
Query Match 58.5%; Score 292.4; DB 23; Length 808;
Best Local Similarity 85.2%; Pred. No. 8,6e-92;
Matches 380; Conservative 0; Mismatches 46; Indels 20; Gaps 4;
XX
66 ACTGAGAGACGAGCTAGCTGATTTCTTAGGCGCACTTAAGATCCCTAAGCTTAGCTG 125
DB 6 ATTAAAGACGAGACGACCTGATGTCTTAGGCTGATTAAGATCCCTAAGCTTAGCTG 65
XX 126 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGCACCTAGCTCACTTACCAAT 185
DB 66 GAAAGTACGACGCTCACTTTAAACAGCGGGCTTGCACCTAGCTCACTTACCAAT 125
XX 186 C-----AGAGAGCTCACTTAAGCTTAAGTGAAGACAGAGGTTAAAGAAATA 236
DB 126 CAGGTAGTAAAGAGAGCTCACTTAAGGCTTAAGGCAAAAGAGAGGTTAAAGAAATA 185
XX 237 GCCAATCATCTATTGCTGAGAGACAGACAGAGAGGACCAACATCGGGATTAAGCCAG 296
DB 186 GCCAATCATCTATTGCTGAGAGACAGACAGAGAGGACCAACATCGGGATTAAGCCAG 245
XX 297 GCATTGAGCTGGCAAGACAGACGCCCCCTTGGGCTCCCTCTTGTGATGAGAGCTGTT 356
DB 246 GCATTGAGCTGGCAAGAGGC-TACCCCTCTGAGGCTCACTCCCTTTGATGAGAGCTCTG 304
XX 357 TTCAATGCTATTCTCACTTAATTAATCTTGCACATG--TACTCTTGTGCTCATGTTCTT 414
DB 305 T-----TTTCACTCTATTAAATCTTGCACATGCACTCTTGTGCTCATGTTCTT 356
XX 415 ACGGCTGAGCTGAGCTTTTGTGCTACCGTCCACGACTCTGTGTTGCCACCGAGACC 474
DB 357 AAGCTTGAAGCTGAGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 416
XX 475 TGCCGCTGACTCCCATCCCTCTGAGT 500
DB 417 CACGCTGACTTCCACCCCTCCGAGT 442
XX
RESULT 14
AAS84189/c
ID AAS84189 standard; cDNA; 1243 BP.
XX
AC AAS84189;
XX

DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #19993.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
XX
XX P-PDB; ABG20002.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 1; SEQ ID No 19993; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 1243 BP; 285 A; 334 C; 358 G; 266 T; 0 other;
SQ
Query Match 58.2%; Score 290.8; DB 23; Length 1243;
Best Local Similarity 85.0%; Pred. No. 4e-91;
Matches 379; Conservative 0; Mismatches 47; Indels 20; Gaps 4;
QY 66 ACTGAGAGACGAGCTAGTGGATTTCTAGGCGCAGTAAGATCCCTAGCCTAGCTGG 125
DB 1238 ATTGAAGAGACGAGCAGCTGTGAGTGTCTAGGCTGACTTAAGATCCCTAGCCTAGCTGG 1179
QY 126 GAAAGTGAACCGCTCCACCTTTAAACAGGGGGCTTGCACCTAGCTACACCTGACCAAT 185
DB 1178 GAAAGTGAACCGCTCCACCTTTAAACAGGGGGCTTGCACCTAGCTACACCTGACCAAT 1119
QY 186 C-----AGAGAGCTCACTAAATGCTAAATTAGGCAAGAGGAGGAAGAGATA 236
DB 1118 CAGGTAGTAAGAGAGCTCACTTAAGAGCTTAATTAGGCAAGAGGAGGAAGAGATA 1059
QY 237 GCCAATCATCTATTGCTGAGAGCAGCAGAGGAGCAACCAATCGGAGATATAAACCAG 296

DB 1058 GCCAATCATCTATTGCTGAGAGCAGCAGGAGGAGGAGCAATGACCGGATTTAAACCAG 999
QY 297 GCATTCGAGCTGGCAACAGAGCCCTTTGGGTCCCTTGGCTCTTGTATGGAGCTGT 356
DB 998 GCATTCGAGCTGGCAACAGG-C-TACCTCTCTGGGTACCTCTTGTATGGAGCTGTG 940
QY 357 TTCATGCTATTTCACCTATTAATCTTGCACAGT--CACTCTTGTGTCATGTTCTT 414
DB 939 T-----TTTCACTCTATTAATCTTGCACAGCACTTTCTGTGTCGATTTGTT 888
QY 415 ACGGCTGAGCTGAGCTTTTGTCTCACCGTCCACCACTGCTGTTTGCACCAACCGAGACC 474
DB 887 AGGCTCTGAGCTGAGCTGTTGCTGTCTGTCACCACTGCTGTTGCTGCGTCGACAGCC 828
QY 475 TGGCGTGAAGCTCCATCCCTCTGAT 500
DB 827 CGCAGCTGACTCTTACCCCTCCGAT 802
RESULT 15
AAK87491
ID AAK87491 standard; DNA; 2629 BP.
XX
XX AAK87491;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:42303.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX MO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JUN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 15-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218390.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225477.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.

PR	22-AUG-2000	2000US-0226868
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227109
PR	30-SEP-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0228509
PR	05-SEP-2000	2000US-0228513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231245
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231415
PR	08-SEP-2000	2000US-0233080
PR	12-SEP-2000	2000US-0233081
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233066
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234224
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235835
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	02-OCT-2000	2000US-0237935
PR	13-OCT-2000	2000US-0239370
PR	13-OCT-2000	2000US-0239371
PR	20-OCT-2000	2000US-0241960
PR	20-OCT-2000	2000US-0241961
PR	20-OCT-2000	2000US-0241962
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241788
PR	20-OCT-2000	2000US-0241807
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241810
PR	20-OCT-2000	2000US-0244617
PR	01-NOV-2000	2000US-0244618
PR	01-NOV-2000	2000US-0246574
PR	08-NOV-2000	2000US-0246575
PR	08-NOV-2000	2000US-0246576
PR	08-NOV-2000	2000US-0246577
PR	08-NOV-2000	2000US-0246578
PR	08-NOV-2000	2000US-0246579
PR	08-NOV-2000	2000US-0246580
PR	08-NOV-2000	2000US-0246581
PR	08-NOV-2000	2000US-0246582
PR	08-NOV-2000	2000US-0246583
PR	08-NOV-2000	2000US-0246584
PR	08-NOV-2000	2000US-0246585
PR	08-NOV-2000	2000US-0246586
PR	08-NOV-2000	2000US-0246587
PR	08-NOV-2000	2000US-0246588
PR	08-NOV-2000	2000US-0246589
PR	08-NOV-2000	2000US-0246590
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207

PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCT INC.
 PA
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI, 2001-483426/52.
 DR
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Disclosure; SEQ ID NO 42303; 3071bp + Sequence Listing: English.
 XX
 XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK81699 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 2629 BP; 763 A; 616 C; 587 G; 663 T; 0 other;

Query Match	54.5%	Score 287.6	DB 22	Length 2629
Best Local Similarity	87.5%	Pred. No. 8,38-90		
Matches	377	Conservative	0	Mismatches 49
			Indels	20
			Gaps	4
QY	66	ACTGAGACAGAGACTAGTGGATTTCCTAGGCGCACTAAGAAATCCCTAAGCTGCTGG	125	
Db	81	ACTGAGAGACAGAGACTAGTGGATTTCCTAGGCTGCTAAGATCCCTAAGCTTAGCTGG	140	
QY	126	GAGGTGACACAGCTCCACTTTTAAACAGCGGGCTTGCAACTTAGCTCAACCTGACCAAT	185	

```
Db 141 GAAATTGACCAAGCTCCCTTTAAACAGGGGCTTGCAATTAGTCACACCCGACCAAT 200
QY 186 C-----AGAGGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATA 236
Db 201 CAGGTACTAAAGAGGCTCACTAAATGCTAATTAGGAAACAGAGGTAAAGAAATA 260
QY 237 GCCAATCATCTAATTGCTGAGAGCAGACAGAGGGAACAATCGGATATAAACCCAG 296
Db 261 GCCAATCATCTAATGCTGAGAGCAGACAGAGGGAACAATGATCAAGATATAAACCCAG 320
QY 297 GCATTGAGCTGGCAAGACAGACCCCTTGGGTCCTTCCCTTGTATGGAAGCTGTT 356
Db 321 GCATTGAGCTGGCAAGAGGAGC-TACCTCTTGGGGTCCCTCCCTTGTATGGAAGCTCTG 379
QY 357 TTCATGCTATTTGACCTAATTAAATCTTGAACG--CACCTCTGAGTCATGTTCTT 414
Db 380 T-----TTTCACTAATTAAATCTTGAATGCAACACTTTCGTGATAGTGTGTC 431
QY 415 ACGGCTGAGCTGAGCTTTTGTCTCACCCTCCACCACTGCTGTTTGCACACCGAGACC 474
Db 432 ACGCTCAAGCTGAGCTTTGCTCACCCTCCACCACTGCTGTGTGCGCGCTGTCAAGACC 491
QY 475 TCCCGCTGACTCCCATCCCTCTGGAT 500
Db 492 CACAGCTGACTTCATCCCTCTGGAT 517
```

Search completed: April 17, 2003, 05:06:35
Job time : 198.426 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:35:35 ; Search time 34.6205 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500
Perfect score: 500
Sequence: 1 cccctg9g9c9g9cctccctt.....tgaccatccatccctcgat 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338361 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	284.8	57.0	2946	4	US-09-175-928-3	Sequence 3, Appl1
2	200	40.0	279	1	US-08-886-878A-50	Sequence 50, Appl1
3	200	40.0	279	1	US-08-721-489-4	Sequence 4, Appl1
4	44	8.8	7218	1	US-08-232-463-14	Sequence 14, Appl1
5	35.4	7.1	7218	1	US-08-232-463-14	Sequence 14, Appl1
6	30.8	6.2	9763	3	US-08-973-273-1	Sequence 1, Appl1
7	30.6	6.1	1078	2	US-08-933-750C-90	Sequence 90, Appl1
8	30.6	6.1	1078	2	US-09-234-613-90	Sequence 90, Appl1
9	30.6	6.1	11827	4	US-09-739-455-3	Sequence 3, Appl1
10	29.8	6.0	16389	4	US-09-741-154-3	Sequence 3, Appl1
11	29	5.8	2023	3	US-08-961-083-199	Sequence 199, App
12	29	5.8	32768	4	US-08-961-527-71	Sequence 71, Appl1
13	29	5.8	36651	4	US-09-738-894A-3	Sequence 3, Appl1
14	28.8	5.8	3386	1	US-08-703-809-2	Sequence 2, Appl1
15	28.8	5.8	3386	1	US-08-703-808-2	Sequence 2, Appl1
16	28.8	5.8	3386	2	US-08-914-066-2	Sequence 2, Appl1
17	28.8	5.8	3386	2	US-08-914-066-2	Sequence 2, Appl1
18	28.8	5.8	3386	2	US-08-747-108A-2	Sequence 2, Appl1
19	28.8	5.8	3386	3	US-09-211-631-2	Sequence 2, Appl1
20	28.8	5.8	3386	3	US-09-265-628-2	Sequence 2, Appl1
21	28.8	5.8	3386	4	US-09-001-141-2	Sequence 2, Appl1
22	28.8	5.8	3386	4	US-09-532-803-2	Sequence 2, Appl1
23	28.6	5.7	2646	4	US-09-221-017B-558	Sequence 558, App
24	28.4	5.7	2897	4	US-09-066-046-38	Sequence 38, Appl1
25	28.4	5.7	2897	5	PCT-US96-05320A-637	Sequence 637, App
26	28	5.6	1262	2	US-08-989-386-4	Sequence 4, Appl1
27	27.8	5.6	15567	4	US-09-627-376-3	Sequence 3, Appl1

28	27.6	5.5	974	1	US-08-848-932-3	Sequence 3, Appl1
29	27.6	5.5	974	2	US-09-008-180-3	Sequence 3, Appl1
30	27.6	5.5	1017	4	US-09-330-611-5	Sequence 5, Appl1
31	27.6	5.5	1666	1	US-08-848-932-1	Sequence 1, Appl1
32	27.6	5.5	1666	2	US-09-008-180-1	Sequence 1, Appl1
33	27.6	5.5	2886	2	US-08-687-080-55	Sequence 55, Appl1
34	27.4	5.5	1631	1	US-08-468-853-5	Sequence 5, Appl1
35	27.4	5.5	1631	1	US-08-468-853-5	Sequence 5, Appl1
36	27.4	5.5	1631	1	US-08-310-357-5	Sequence 5, Appl1
37	27.4	5.5	1631	1	US-08-468-852-5	Sequence 5, Appl1
38	27.4	5.5	1631	2	US-08-468-857-5	Sequence 5, Appl1
39	27.4	5.5	1741	4	US-09-232-160-4	Sequence 4, Appl1
40	27.4	5.5	3054	4	US-09-484-970B-138	Sequence 138, App
41	27.4	5.5	5761	1	US-07-749-001-2	Sequence 2, Appl1
42	27.4	5.5	5761	1	US-08-154-198-2	Sequence 2, Appl1
43	27.4	5.5	5761	1	US-08-463-335-2	Sequence 2, Appl1
44	27.4	5.5	5761	2	US-08-464-022A-2	Sequence 2, Appl1
45	27.2	5.4	360	4	US-09-171-945-54	Sequence 54, Appl1

ALIGNMENTS

RESULT 1
US-09-175-928-3
Sequence 3, Application US/09175928A
Patent No. 6312921

GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Ma, Sha
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: 6006B AJ172A
CURRENT APPLICATION NUMBER: US/09/175, 928A
CURRENT FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-09-175-928-3

Query Match 57.0%; Score 284.8; DB 4; Length 2946;
Best Local Similarity 95.5%; Pred. No. 5.1e-89;
Matches 315; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY	66	ACTGAGAGACAGACAGTACGATTTCTTAGGCCGACATGAATCCCTAAGCCTAGCTGG	125
DB	2601	ACTGAGAGACAGACAGTACGATTTCTTAGGCCGACATGAATCCCTAAGCCTAGCTGG	2660
QY	126	GAAGGTACACAGCTTCACTTTAAACAGCGGGCTTGGCAATTAGCTCACCTGACCAAT	185
DB	2661	GAAGGTACACACATTCACCTTTAAACAGCGGGCTTGCATTTAGCTCACCTGACCAAT	2720
QY	186	CAGAGAGCTACATTAATGCTAATTAGGCAAGACAGAGTAAAGTAATGACCAATCAT	245
DB	2721	CAGAGAGCTACATTAATGCTAATTAGGCAAGACAGAGTAAAGTAATGACCAATCAT	2780
QY	246	CTATTGCTTAGAGACACAGACAGAGGACACAAATCGGATATTAACCCAGGATTCGAG	305
DB	2781	CTATTGCTTAGAGACACAGACAGAGGACACAAATCGGATATTAACCCAGGATTCGAG	2840
QY	306	CTGGCAACAGACACCCCTTTGGGTCCCTTTGGTATGGAG--CTGTTTCATGC	363
DB	2841	CTGGCAACAGGCA-ACCCCTTTGGGTCCCTTTGGTATGGAGCTGTGTTTATGC	2899

QY 364 TATTCACTCTATTAATCTTGCAACTGCA :93
| | | | | | | | | | | | | | | | | |
Db 2900 TATTTCACTCTATTAATCTTGCAACTGCA 2929

RESULT 2

```

US-08-686-878A-50
: Sequence 50, Application US/08686878A
: Patent No. 5708157
:
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavoille, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Evans, Cheryl
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 71
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/686,878A
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 50:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 279 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
:
US-08-686-878A-50

```

Query Match	40.0%;	Score 200;	IB 1;	Length 279;
Best Local Similarity	88.2%;	Pred. No. 4, fe-60;		
Matches 232; Conservative	1;	Mismatches 27;	Indels 3;	Gaps 2

QY	133	ACAGCTGCACCTTTAAACAGGGGGCTTGCACCTTAGCTCACCTACCACTAGAGAG	129
Ds	1	RCACATTCACCTTTAAACAAGGGGNTTGCAAANBAGATNACCTTAGCCACATCAGAG	60
QY	193	CTCACTAAATAGCTATTATTAGGCAAAGACAGAGGTTAAGAAATNAGCCATCATCTATTGC	252
Ds	61	NTCANTAAATGATNATTNTGGCMAAAACAGGAGGTANAGAAATAGCCATCATCTATTGC	120
QY	253	CTGAGAGCACAGCAGAGAGGACAAACATTCGGGATTAAACCCAGCATTCGAGCTGGCAA	312
Ds	121	CTGAGAGCACAGCAGAGAGGACAAAGATTCGGGATTAAACCCAGTTTNTAGCGCGCAA	180
QY	313	CAGCAGCCCCCTTTGGGCTTCCTTCCTTTGATGGGAC--TGTTTTCATGCTATTTCA	370
Ds	181	CGGCA-ACCCCCCTTTGGGCTTCCTTCCTTTGATGGGACCTNTGTTTCAAGCTATTTCA	238
QY	371	CTCTATTAAATCTTGCAACTGCA	393

Db 240 NTNTATTAATNTGCACTGCA 262

RESULT 3

```

US-08-721-489-4
: Sequence 4, Application US/08721489
: Patent No. 5785465
:
: GENERAL INFORMATION:
:   APPLICANT: Jacobs, Kenneth
:   APPLICANT: McCoy, John
:   APPLICANT: Lavallee, Edward
:   APPLICANT: Racie, Lisa
:   APPLICANT: Merberg, David
:   APPLICANT: Treacy, Maurice
:   APPLICANT: Spaulding, Vikki
:   TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
:   TITLE OF INVENTION: ENCODING THEM
:   NUMBER OF SEQUENCES: 5
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Genetics Institute, Inc.
:   STREET: 87 CambridgePark Drive
:   CITY: Cambridge
:   STATE: Massachusetts
:   COUNTRY: U.S.A.
:   ZIP: 02140
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent in Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/721,489
:   FILING DATE:
:   CLASSIFICATION: 530
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Brown, Scott A.
:   REGISTRATION NUMBER: 32,724
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (617) 498-8224
:   TELEFAX: (617) 876-5851
:   INFORMATION FOR SEQ ID NO: 4:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 279 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: double
:     TOPOLOGY: linear
:   MOLECULE TYPE: CDNA
:
US-08-721-489-4

```

Query Match	40.0%;	Score 200;	DB 1;	Length 279;
Best Local Similarity	88.2%;	Pred. No. 4.8e-60;		
Matches 232; Conservative	1;	Mismatches 27;	Indels 3;	Gaps 2;

Oy	133	ACAGGTCCACCTTTAAACACGGGGCTTGCAACTAGCTACACCTGACCAATCCAGAG	192
	:		
Db	1	RCACATCCACCTTTAAACAGGGGNTTGCAAAATAATNACCTTACCAATCAGAG	60
Oy	193	CTCACTAAATGCTATTAATGAGCAAGA CAGAGGTATAAGAAATAGCCATCATCTATTGC	252
Db	61	NTCANTAAATGATNATTNTGCAAAAA CAGAGGTTAAAGAAATAGCCATCATCTATTGC	120
Oy	253	CTGAGAGACAGCAGAGAGGACCAACATCGGATATATAACCCAGGGCACTTGAGCTGGCA	312
Db	121	CTGAGAGACACGAGAGGACCAATGATCGGATATATAACCCAGTTTNTGAGCCGGCA	180
Oy	313	CAGCAGCCCCCTTTGGGTCCCTTCCCTTGTGATGGAGC--TGTTTTCATGCTATTCA	370
Db	181	CGGCA-ACCCCTTTGGGTCCCTTCCCTTGTATGGAGCGTNTGTTTCATGCTATTCA	239
Oy	371	CTCATTAATCTTGCAACTGCA	393
	:		
Db	240	NTNATTAATNTTGCAACTGCA	262


```

RESULT 4
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpc-F18
US-08-232-463-14

Query Match 8.8%; Score 44; DB 1; Length 7218;
Best Local Similarity 4.1%; Pied.No. 7,4e-05;
Matches 8; Conservative 123; Mismatches 63; Indels 0; Gaps 0;

QY 303 GAGCTGGGAACGACGAGCCCCCTTTGGGTCCCTTCCTTGATGAGAGCTGTTTCANG 362
Db 1056 GAGCTGGGATATATATCTTGACACTGCACTCTTGTCACATGTTTCTACGGCTCG 1115
QY 363 CTATTTACTCTATTAATCTTGACACTGCACTCTTGTCACATGTTTCTACGGCTCG 422
Db 1116 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1175
QY 423 AGCTGAGCTTTTGCTCACCGTCACACACTGCTGTTTGCCACACGCGAGACTGCGCGTG 482
Db 1176 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1235
QY 483 ACTCCATCCCTCT 496
Db 1236 YYYYYYYYYYYY 1249

RESULT 5

```

```

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpE-Fls
US-08-232-463-14

Query Match          7.1%; Score 35.4; DB 1; Length 7218;
Best Local Similarity 3.0%; Pred.No. 0.074; Mismatches 125; Indels 0; Gaps 0;
Matches 9; Conservative 169;

      23 GGAGTGAAGGAAAAACGGCTGAGATACAGAATATCTTGCAACTGAGACAGACTA 82
Dbb   : ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      1335 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296
Qy    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      83 GCTGATTTCCTAGGCCCATGAATCCCTTAAGCTTAGCTGGAAAGGTGACCAAGTCCA 142
Db    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236
Qy    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      143 CCTTTAAACAGCGGCTTGCAACTTAGCTCACACTGCATCGCAATCAGAGAGCTCAATAAA 202
Db    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176
Qy    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      203 TGCTAATTAGCAAAGACAGAGCGTAAGAAATAGCCATCATCTAATTGCTCGTAGAGAC 262
Db    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116
Qy    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      263 AGCAGAGAGGACAACAATCGGATATAAACCCAGGCAATTCGAGCTGGACAAGACCCC 322
Db    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCTC 1056
Qy    : ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
      323 CCT 325

```

Db 1055 CCT 1053

```

|||
US-08-973-273-1/c
; Sequence 1, Application US/08973273
; Patent No. 6140085
; GENERAL INFORMATION:
; APPLICANT: Dean, Caroline
; APPLICANT: MacKnight, Richard C
; APPLICANT: Bancroft, Ian
; APPLICANT: Lister, Clare K
; TITLE OF INVENTION: Genetic Control of Flowering
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 6140085th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,273
; FILING DATE: 01-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01332
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9511196.9
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ms Mary J Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-29
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9763 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Arabidopsis thaliana
; STRAIN: Columbia
; US-08-973-273-1

Query Match 6.2%; Score 30.8; LB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 3.5;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 176 CCTGACCATCGAGAGCTCACTAAATGCTATTAGCGAAGACAGAGGTAAGAAT 235
Db 4845 CCTTGAAGAAGAGACTATATCTGATAGTAAGCAAGAAACGGAAGAAAA 4786

QY 236 AGCCATCATCTAT 249
Db 4785 GGATCAACATCAT 4772

RESULT 7
US-08-933-750C-90
; Sequence 90, Application US/08933750C
```

```

; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shan, Fuyi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 90:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SINITOT01
; CLONE: 2184712
; US-08-933-750C-90

Query Match 6.1%; Score 30.6; DB 2; Length 1078;
Best Local Similarity 58.3%; Pred. No. 1.2;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGGCAATTCAGCTGCGAAGACAGCCCTTTGGTCCCTTTGATG 347
Db 941 TCAACTGGAAGAAGATCTTGCTACAGAGAGAGCCCTTGCGCTCCCTTCCTT---TG 996

QY 348 GGAGCTGTTTCATGCTATTCTATTAAATCTTGCAAGTCGACCTCTGTGCAT 407
Db 997 ATAGCAGTTATTAATGCTTGTTCCTCAATTAATCTGGGCAAGATGAATCCATGTCAT 1056

QY 408 GTTCTT 414
Db 1057 ACTGCCT 1063

RESULT 8
US-09-234-613-90
; Sequence 90, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
```

```
/ APPLICANT: Hillman, Jennifer L.
/ APPLICANT: Bandman, Olga
/ APPLICANT: Shah, Puri
/ APPLICANT: Au-Young, Janice
/ APPLICANT: Yue, Henry
/ APPLICANT: Guegler, Karl J.
/ APPLICANT: Corley, Neil C.
/ TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
/ NUMBER OF SEQUENCES: 98
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Incyte Pharmaceuticals, Inc.
/ STREET: 3174 Porter Drive
/ CITY: Palo Alto
/ STATE: CA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/234,613
/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/933,750
/ FILING DATE: September 23, 1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Billings, Lucy J.
/ REGISTRATION NUMBER: 36,749
/ REFERENCE/DOCKET NUMBER: PF-0356 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-855-0555
/ TELEFAX: 415-845-4166
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 90:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1078 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: SININOT01
/ CLONE: 2184712
/ US-09-234-613-90

Query Match
Best Local Similarity 6.1%; Score 30.6; DB 3; Length 1078;
Matches 74; Conservative 0; Mismatches 49; Indels 4; Gaps 1;

QY 288 TAAACCCAGCATTGAGCTGCGAAGACGACCCCTTTGGTCCCTTGGTATG 347
DB 941 TCAACTGTGAAGAGATCTTGTACAGAGAGACCTTGGGCTCCCTTCTT---TG 996
QY 348 GGAAGCTGTTTTCATGCTATTTCATTTAATCTTGAACGACGACCTTGGTCCAT 407
DB 997 ATGACGATTATATGCTTCTTGTCCCAATAAAGCGGACAGATGAATCTAGTCTAT 1056
QY 408 GTTTCCTT 414
DB 1057 ACTGCCT 1063

RESULT 9
US-09-739-455-3
/ Sequence 3, Application US/09739455
/ Patent No. 6413756
/ GENERAL INFORMATION:
/ APPLICANT: YAN, Chunhua et al
/ TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
/ TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
/ TITLE OF INVENTION: THEREOF
```

```
/ FILE REFERENCE: CL000653
/ CURRENT APPLICATION NUMBER: US/09/739,455
/ CURRENT FILING DATE: 2000-12-19
/ NUMBER OF SEQ ID NOS: 23
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 11827
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(11827)
/ OTHER INFORMATION: n = A,T,C or G
/ US-09-739-455-3

Query Match
Best Local Similarity 6.1%; Score 30.6; DB 4; Length 11827;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 117 CTTAGCTGGAGAGGTGACGCTCCACCTTTAAACAGGGGCTTGCACTAGCTCAC 176
DB 10853 CCAGCTGCTGAGAGTGGAGACTGCTGCTTTGTTGGCGCTTGTCTTAATCAGTT 10912
QY 177 CTGACCAATCAGAGAGCTCACTAAAAATGCTAATTAGGCAAGAAGAGGTAAAGAATA 236
DB 10913 CCTCTTAGATTATTATACCTAAAAAAATTAAGTTTGAAGAAATAGAGATA 10972
QY 237 GCCATCATCTATTGCTGANG 257
DB 10973 CAGAAACATGATTTCCAGCG 10993

RESULT 10
US-09-741-154-3/C
/ Sequence 3, Application US/09741154
/ Patent No. 6437110
/ GENERAL INFORMATION:
/ APPLICANT: BEASLEY, Ellen M, et al
/ TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
/ TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
/ FILE REFERENCE: CL001061
/ CURRENT APPLICATION NUMBER: US/09/741,154
/ CURRENT FILING DATE: 2000-12-21
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 16389
/ TYPE: DNA
/ ORGANISM: Human
/ US-09-741-154-3

Query Match
Best Local Similarity 6.0%; Score 29.8; DB 4; Length 16389;
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 74 ACAGACTAGCTGATTCTCTAGCGGACGACTAAGAACCTTAAGCTTGGAGGATGA 133
DB 1196 ACAGACTAATGTGAATGCTGAGTCAATTAATTCCTGATCCGACGCGGAGAGGG 1137
QY 134 CCACGTCCACCTT 146
DB 1136 ATGCTCTGAGCTT 1124

RESULT 11
US-08-961-083-199
/ Sequence 199, Application US/08961083
/ Patent No. 6159469
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
/ NUMBER OF SEQUENCES: 452
```

```

CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 2023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-083-199

Query Match
Best Local Similarity 5.8%; Score 29; DB 3; Length 2023;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTAATTAGCAAGACAGAGGTAAAGAAATACCATCATCTATTGCTGAGAGCAC 262
DB 1711 TGATATTGAGTTTAAGATAGCAGCTAGAGTAAATAGTTATGATCATTTGGCTCAGATGAC 1770

QY 263 A 263
DB 1771 A 1771

RESULT 12
US-08-961-527-71
Sequence 71, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunach
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
```

```

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-71

Query Match
Best Local Similarity 5.8%; Score 29; DB 4; Length 32768;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 203 TGCTAATTAGCAAGACAGAGGTAAAGAAATACCATCATCTATTGCTGAGAGCAC 262
DB 7947 TGATATTGAGTTTAAGATAGCAGCTAGAGTAAATAGTTAGTATGATGCTGCTCAGATGAC 8006

QY 263 A 263
DB 8007 A 8007

RESULT 13
US-09-738-894A-3/C
Sequence 3, Application US/09738894A
Patent No. 6331423
GENERAL INFORMATION:
APPLICANT: GUGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL000636
CURRENT APPLICATION NUMBER: US/09/738,894A
CURRENT FILING DATE: 2000-12-18
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 36651
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(36651)
OTHER INFORMATION: n = A,T,C or G
US-09-738-894A-3

Query Match
Best Local Similarity 5.8%; Score 29; DB 4; Length 36651;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 15 TCCTTTCTGGATGAGGCAAGACGCTGAGATACGCAATTATCTTGCACATGAGAGA 74
DB 6891 TCCTACCTGGAACATGAGCCCTGATGGCTGGGGCTACAGCACCATCTTGCAGCATGGA 6832

QY 75 CAGGACTAG 83
DB 6831 AAAGGCTAG 6823

RESULT 14
US-08-703-809-2
Sequence 2, Application US/08703809
Patent No. 5716808
GENERAL INFORMATION:
APPLICANT: Raymond, Christopher K.
TITLE OF INVENTION: GENETIC ENGINEERING OF PICHIA METHANOLICA
```

```

;
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,809
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 96-18
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-703-809-2
;
;
; Query Match
; Best Local Similarity 49.3%; Score 28.8; DB 1; Length 3386;
; Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
;
; QY 307 TGGCAACAGCAGCCCCCTTGGGTCCTTCCTTTGTATGGAGCTGTTTCAATGCTAT 366
;      |||||
; Db 856 TGGCAACAGATAGTATCTATCTGCTTAATTCATCCACTTGGGAACGGCTCTCTTTAC 915
;
; QY 367 TTGACTATTAATCTTGAACAGCACTCTTCTGGTCCAGTTTCTTAAGGCTCGAGCT 426
;      |||||
; Db 916 CCCAGATTCTCAAGCTAATATCTGCCCTTGTCTATTTGCTTTCCTGTAACAACGG 975
;
; QY 427 GAGCTTTGCTACCGTCACCACTGCTGTTT 458
;      |||||
; Db 976 GAGCTTTGCTCCCATCTCTTGTCTTTT 1007
;
;
; RESULT 15
; US-08-703-808-2
; Sequence 2, Application US/08703808
; Patent No. 5736383
; GENERAL INFORMATION:
; APPLICANT: Raymond, Christopher K.
; TITLE OF INVENTION: PREPARATION OF PICHIA METHANOLICA AUXOTROPHIC
; TITLE OF INVENTION: MUTANTS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
;

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703,808
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, Gary E
; REGISTRATION NUMBER: 31-648
; REFERENCE/DOCKET NUMBER: 96-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6673
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; US-08-703-808-2
;
;
; Query Match
; Best Local Similarity 49.3%; Score 28.8; DB 1; Length 3386;
; Matches 75; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
;
; QY 307 TGGCAACAGCAGCCCCCTTGGGTCCTTCCTTTGTATGGAGCTGTTTCAATGCTAT 366
;      |||||
; Db 856 TGGCAACAGATAGTATCTGCTTAATTCATCCACTTGGGAACGGCTCTCTTTAC 915
;
; QY 367 TTGACTATTAATCTTGAACAGCACTCTTCTGGTCCAGTTTCTTAAGGCTCGAGCT 426
;      |||||
; Db 916 CCCAGATTCTCAAGCTAATATCTGCCCTTGTCTATTTGCTTTCCTGTAACAACGG 975
;
; QY 427 GAGCTTTGCTACCGTCACCACTGCTGTTT 458
;      |||||
; Db 976 GAGCTTTGCTCCCATCTCTTGTCTTTT 1007
;
;
; Search completed: April 17, 2003, 07:37:57
; Job time : 73.6205 secs
;

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:57:20 ; Search time 123.169 Seconds
(without alignments)
4077.979 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500
Perfect score: 500
Sequence: 1 cccctggggcggccttcctt.....tgactccatccctcgtat 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published_Applications_NA.*
2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
3: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	348.2	69.6	1329	7 US-08-979-847-108	Sequence 108, App
2	301	60.2	180557	12 US-10-003-806-6	Sequence 6, Appl1
3	301	60.2	180557	12 US-10-003-806-9	Sequence 134, App
4	284.8	57.0	2946	9 US-10-114-893-134	Sequence 1, Appl1
5	283.8	56.8	2930	10 US-09-502-535-1	Sequence 102, App
6	270.4	54.1	635	7 US-08-979-847-102	Sequence 538, App
7	261.4	52.3	410	10 US-09-880-107-538	Sequence 8173, App
8	259.2	51.8	541	10 US-09-864-761-8173	Sequence 3, Appl1
9	232.2	46.4	336014	10 US-09-731-231A-3	Sequence 4444, App
10	226.2	45.2	1894	9 US-09-864-761-4444	Sequence 1654, App
11	218	43.6	15425	9 US-10-091-504-1654	Sequence 14951, A
12	218	43.6	15425	10 US-09-764-869-1654	Sequence 50, Appl1
13	204.6	40.9	569	10 US-09-864-761-14951	Sequence 20462, A
14	200	40.0	279	12 US-10-040-916-50	Sequence 895, App
15	126	25.2	246	10 US-09-864-761-20462	Sequence 1, Appl1
16	118.2	23.6	446	10 US-09-811-284-3	Sequence 895, App
17	111.6	22.3	475	10 US-09-864-761-895	Sequence 1, Appl1
18	109.4	21.9	3524	10 US-09-972-724-1	Sequence 4153, App
19	102	20.4	409	10 US-09-864-761-4153	

C 20	101	20.2	559	10	US-09-864-761-7501	Sequence 7501, App
C 21	97.8	19.6	504	10	US-09-864-761-7027	Sequence 7027, App
C 22	93	18.6	440	10	US-09-864-761-3694	Sequence 3694, App
C 23	79.2	15.8	579	10	US-09-864-761-13678	Sequence 13678, A
C 24	50.4	10.1	88	10	US-09-864-761-20907	Sequence 20907, A
C 25	49.8	10.0	525	10	US-09-893-737-31	Sequence 31, Appl1
C 26	48.8	9.8	367	10	US-09-864-761-30194	Sequence 30194, A
C 27	37.8	7.6	275	10	US-09-864-761-17675	Sequence 17675, A
C 28	34.6	7.0	1508	9	US-09-735-713A-7	Sequence 7, Appl1
C 29	34.6	6.9	79	10	US-09-864-761-31488	Sequence 31488, A
C 30	32.6	6.5	384	10	US-09-783-590-10544	Sequence 10544, A
C 31	32.6	6.5	653	9	US-10-184-644-402	Sequence 402, App
C 32	32.6	6.5	653	9	US-10-184-634-402	Sequence 402, App
C 33	31.8	6.4	468	9	US-09-796-692-4666	Sequence 4666, App
C 34	31.2	6.2	473	10	US-09-864-761-11001	Sequence 11001, A
C 35	31.2	6.2	10514	10	US-09-764-877-3470	Sequence 3470, App
C 36	31.2	6.2	197997	10	US-09-822-246-3	Sequence 3, Appl1
C 37	31.1	6.2	407	10	US-09-864-761-20523	Sequence 20523, A
C 38	30.8	6.2	2000	9	US-09-938-842A-2796	Sequence 2796, App
C 39	30.6	6.1	410	10	US-09-867-701-4129	Sequence 4129, App
C 40	30.6	6.1	1078	10	US-09-840-787-90	Sequence 90, Appl1
C 41	30.6	6.1	8220	10	US-09-797-908-3	Sequence 3, Appl1
C 42	30.6	6.1	8522	10	US-09-817-181-3	Sequence 3, Appl1
C 43	30.2	6.0	414	9	US-10-123-155-418	Sequence 418, App
C 44	30.2	6.0	2012	9	US-09-746-783-147	Sequence 147, App
C 45	30.2	6.0	2014	9	US-10-036-041-39	Sequence 39, Appl1

ALIGNMENTS

RESULT 1
US-08-979-847-108
Sequence 108, Application US/08979847
Publication No. US20030039664A1

GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADIL, FLORENCE
APPLICANT: JOLIVET-REYNARD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY

TITLE OF INVENTION: TUBE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 108:

	Query Match	Score	DB	Length
Best Local Similarity	83.2%	Pred. No. 5.3e-94		
Matches 397, Conservative	0	Mismatches 60	Indels 20	Gaps 4
35	AAACGCTGGAATACAGCAATTATTTGCCACTGAGAACAGGAACTTACTGATTTCTT	94		

QY	35	AAAGCCCTGAGATGACCAATTATCTTGCAACTGAGAGACAGACTAGCTGGATTTCCT	94
Db	59022	AAGGACCTGAGTCAATGGAAATGACATCATATGTAGAGACAGGACTAGCTGGATTTCCT	589635
QY	95	AGGCGCACTAAGAAATCCCTTAAGCTTGGAGGTGACCACTTCAACTTTAAACAG	154
Db	58962	AGCGGCACTAAGAAATCCCTTAAGCTTGGAGGTGACCGCTTCAACTTTAAACATG	589030
QY	155	GGGCTTGCAACTTAGCTCAACTTGACCAATC-----AGAGAGCTCACTAAATGC	205
Db	58902	GGGCTTGCAACTTAGCTCAACCCGACCAATCAGATGTAAAGAGAGGTCTCAATAAAATGC	588430
QY	206	TAAATTAGCAAAAGACAGAGGTAAAGAAATAGCCATATCTATTTGCTGAGAGCAAGC	265
Db	58842	TAAATTAGCAAAAGACAGAGGTAAAGAAATAGCCATATCTATTTGCTGAGAGCAAGC	587933

QY	265	AGGAGGAGCAACAAATCGGAGATTTAAACCAAGGCATTCGAGCTGGGCAACAGACGCCCCCT	325
Db	58782	AGGAGGAGCAATGATCAGCATTTAAACCAGGCATCTGAGCCAGCAAGC-TAGCCTT	58724
QY	326	TTGGTCCCTTCCCTTTGTATGAGGAGCTGTTTCAATGCAATTTCACTATTAAATCTTG	385
Db	58723	TTGGTCCCTTCCCTTTGTATGAGGAGCTGTG-----CTTCACTGATTTAAATCTTG	58672
QY	386	CAACTGCA-CTCTTCTGATCATGTTTCTTAAGGCTGAGCTGAGCTTTTGCTACCGT	443
Db	58671	CAGCTGCACTCTCTTTTGATCAACTTTGTCAATGCTTCAGCTGAGCTTCTCTCGCCGT	58612
QY	444	CAACCACTGCTGTTTGCCACACACGCGAGACGCGCGGTACTCCCATTCCTCTGGAT	500
Db	58611	CAACCACTGCTGTTTGCCGCTGTGCGAGACTGCTGCTGACTTCATCCGTCAGAT	58555

```

RESULT 4
US-10-114-893-134
; Sequence 134, Application US/10114893
; Publication No. US20020193567A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallee, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Treacy, David
; APPLICANT: Merberg, Maurice
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Carlin-Duckett, McKeeough
; APPLICANT: Kelleher, Kerry S.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: GI 6000-10A
; CURRENT APPLICATION NUMBER: US/10/114,893
; EARLIER FILING DATE: 2002-04-02
; EARLIER APPLICATION NUMBER: 09/413,232
; EARLIER FILING DATE: 1998-10-06
; NUMBER OF SEQ ID NOS: 321
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 134
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-114-893-134

```

Query Match	57.0%;	Score 284.8;	DB 9;	Length 2946;
Best Local Similarity	95.5%;	Pred. No. 2.6e-89;		
Matches 315; Conservative	0;	Mismatches 12;	Indels 3;	Gaps 2;

Oy	66	ACTGAGGACAGGACCTAGCTGGATTTCTTAAGGCCACTAAGAACTCTTAAGCTTAGCTGG	125
Db	2601	ACTGAGGACAGGACCTAGCTGGATTTCTTAAGCTGACTAAGAACTCTTAAGCTTAGCTGG	2660
Oy	126	GAGGCTGACCAAGCTCACTTTTAAACACGGGGCTTGCAACTTAGCTCACACTGACCAAT	185
Db	2661	GAGGCTGACCAAGCTCACTTTTAAACACGGGGCTTGCAACTTAGCTCACACTGACCAAT	2720
Oy	186	CAGGAGCTCACATAAATGTCTAATTGGGCAAGAACAGGAGGTAAAGAAATGGCCAATAT	245
Db	2721	CAGGAGCTCACATAAATGCTAATTGGGCAAAAACAGGAGGTAAAGAAATAGCCAATAT	2780
Oy	246	CTATTGCTGTGAGCACAGCAGAGGGACAACTGGSATTTAAACCCGAGCATTTGAG	305
Db	2781	CTATTGCTGTGAGCACAGCAGAGGGACAACTATGGGATTTAAACCCAGCTTTGAG	2840
Oy	306	CTGGCAACAGAGAGCCCCCTTTGGGTCCCTCCCTTGTATGGAG--CTGTTTTATGC	363
Db	2841	CCGGCAAGGGA-ACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTATGC	2839
Oy	364	TATTTCACCTATTAAATCTTGCACTGCA	393

Db 2900 TATTTCACTCTATTAATCTTGCACACTGCA 2929

```

RESULT 5
US-09-902-535-1
; Sequence 1, Application US/09902535
Patent No. US20020102530A1
GENERAL INFORMATION:
APPLICANT: Keith, Jr., James C.
APPLICANT: McCoy, John M.
APPLICANT: Mi, Sha
TITLE OF INVENTION: Methods and compositions for diagnosing
TITLE OF INVENTION: and treating preeclampsia and gestational trophoblastic
TITLE OF INVENTION: disorders
FILE REFERENCE: GIN-6006B4
CURRENT APPLICATION NUMBER: US/09/902,535
CURRENT FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: 60/216,657
PRIOR FILING DATE: 2000-07-06
NUMBER OF SEQ. ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (930)...(2546)
US-09-902-535-1

```

Query Match	56.8%;	Score 283.8;	DB 10;	Length 2930;
Best Local Similarity	-95.4%;	Pred. No. 5.8e-89;		
Matches 314;	Conservative 0;	Mismatches 12;	Indels 3;	Gaps 2;

Qy	66	ACTGAGAGACAGGACTTATGCTGGATTTCTTCAAGGCCCACTTAAAGATATCCCTAAGCTGATGCTGG	125
Db	2603	ACTGAGGAGACAGGACTTATGCTGGATTTCTTCAAGCTGACTTAAAGATCCCTAAGCTGATGCTGG	2662
Qy	126	GAAAGTATCCACGTCACCTTTTAAACAGGAGGCTTGCAACTTATGCTCACACTGACCAAT	185
Db	2663	GAAAGTATCCACATCCACTTTTAAACAGGAGGCTTGCAACTTATGCTCACACTGACCAAT	2722
Qy	186	CAGAGACTCACTAAATGCTTAATTATGAGCAAGACAGAGTTAAAGAAATAGCCATCAT	245
Db	2723	CAGAGACTCACTAAATGCTTAATTATGAGCAAGAAACAGAGTTAAAGAAATAGCCATCAT	2782
Qy	246	CTATATGCTTGAGACACAGCAGGAGGACCAACATGGGATTTAAACCCAGGACTTGGAG	305
Db	2783	CTATATGCTTGAGACACAGCAGGAGGACAAATGATGGGATTTAAACCCAGGACTTGGAG	2842
Qy	306	CTGGCAACAGCAGGCCCCCTTTGGGTCCCTTCCCTTTGATATGGAG--CTGTTTCAATGC	363
Db	2843	CCGGCAACAGGCA-ACCCCCCTTTGGGTCCCTCCCTTTGATATGGAGCTGTGTTTCAATGC	2901
Qy	364	TATTTCACTGATTTAAATCTTGACATGCG	392
Db	2902	TATTTCACTGATTTAAATCTTGCACTGCG	2930

```

RESULT 5
US-08-979-847-102
: Sequence 102, Application US/08979847
: Publication No. US2003003964A1
: GENERAL INFORMATION:
: APPLICANT: PERRONT, HERVE
: APPLICANT: BESEME, FREDERIC
: APPLICANT: BEDIN, FREDERIC
: APPLICANT: PARANHOS-BACCALA, GLAUCIL
: APPLICANT: KOMUNIAN-PRADEL, FLORENCI
: APPLICANT: JOLIVET-REYAUD, COLETTE
: APPLICANT: MANDRAND, BERNARD
: APPLICANT: GARCON, JEREMY

```

APPLICANT: TUKU, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIVER & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-102

Query Match 54.1%; Score 270.4; 18 7; Length 635;
Best Local Similarity 92.7%; Pred. No. 1.3e-84;
Matches 306; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGACTAGCTGATTTCTTAGGCGGACATAGATCTTAAAGCTTAGCTGG 125
DB 291 ACTGAGAGACAGAGACTAGCTGATTTCTTAGGCGGACATAGATCTTAAAGCTTAGCTGG 350
QY 126 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
DB 351 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT 410
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGTAAAGAAATAGCAATCAT 245
DB 411 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGTAAAGAAATAGCAATCAT 470
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACAATCGGATATAAACCAGGCAATTCGAG 305
DB 471 CTATTGCTGAGAGACAGACAGAGGAGCAACAATCGGATATAAACCAGGCAATTCGAG 530
QY 306 CTGGCAACAGACAGCCCCCTTTGGGCTCCCTTTGATATGGAG--CTGTTTTCATGC 363
DB 531 CCGGCAACAGGCA-ACCCCCCTTTGGGCTCCCTTTGATATGGAGCTCTGTTTTCATGC 589
QY 364 TATTTCACCTAATTAATCTTGCAACTGCA 393
DB 590 TATTTCACCTAATTAATCTTGCAACTGCA 619

RESULT 7
US-09-880-107-538/c
Sequence 538, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Hockley, Joseph G.

APPLICANT: Scherf, Uwe
TITLE OF INVENTION: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 538
LENGTH: 410
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538

Query Match 52.3%; Score 261.4; DB 10; Length 410;
Best Local Similarity 91.2%; Pred. No. 1.4e-81;
Matches 300; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

QY 66 ACTGAGAGACAGACTAGCTGATTTCTTAGGCGGACATAGATCTTAAAGCTTAGCTGG 125
DB 329 ACTGAGAGACAGAGACTAGCTGATTTCTTAGGCGGACATAGATCTTAAAGCTTAGCTGG 270
QY 126 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT 185
DB 269 GAAGGTGACAGCTCCACTTTAAACAGGGGCTTGCACTTAGCTCACACCTGACCAAT 210
QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCAT 245
DB 209 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCAT 150
QY 246 CTATTGCTGAGAGACAGACAGAGGAGCAACAATCGGATATAAACCAGGCAATTCGA 304
DB 149 CTATTGCTGAGAGACAGACAGAGGAGCAACAATCGGATATAAACCAGGCAATTCGA 90
QY 305 GCTGGCAACAGACAGCCCCCTTTGGGCTCCCTTTGATATGGAG--CTGTTTTCATGC 362
DB 89 GCGAGCAACAGGCAACCGCCCTTTGGGCTCCCTTTGATATGGAGCTCTGTTTTCATGC 30
QY 363 TATTTCACCTAATTAATCTTGCAACTG 391
DB 29 TATTTCACCTAATTAATCTTGCAACTG 1

RESULT 8
US-09-864-761-8173/c
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27


```

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine ver. 1.1
SEQ ID NO 4444
LENGTH: 1894
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ACO02346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HEBL100, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444

Query Match 45.2%; Score 226.2; DB 10; Length 1894;
Best Local Similarity 86.4%; Pred. No. 8.2e-69;
Matches 291; Conservative 0; Mismatches 28; Indels 18; Gaps 3;

```

```

Qy 237 GCCAATCATCTATTGCTGAGACACAGAGGACACAACTCGGATATTAACCCAG 296
Db 1687 GCCAATCATCTGTTGCTGACAGACACAGAGGACAAATCGGATATTAACCCAG 1746
Qy 297 GCATTGAGCTGGACACAGACAGCCCCCTTTGGGTCCTTCCCTTTGATGGAGCTGT 356
Db 1747 GCATTGAGCTGGACACAGCTACAGC-TACCCCTTTGGGTCCTTCCCTTTGATGGAGCTGT 1805
Qy 357 TTCATGCTATTTCACCTATTAAATCTTGCACTGCA 393
Db 1806 T-----CTTCACTCTATTAAATCTTGCACTGCA 1834

RESULT 11
US-10-091-504-1654/c
Sequence 1654, Application US/10091504
Publication No. US20030059808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1654

Query Match 43.6%; Score 218; DB 9; Length 15425;
Best Local Similarity 85.5%; Pred. No. 2e-65;
Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

```

NUMBER OF SEQ ID NOS: 2442
 SOFTWARE: Patentlin Ver. 2.0
 SEQ ID NO: 1654
 LENGTH: 15425
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-764-869-1654

Query Match 43.6%; Score 218; DB 10; Length 15425;
 Best Local Similarity 85.5%; Pred. No. 2e-65;
 Matches 284; Conservative 0; Mismatches 30; Indels 18; Gaps 3;

QY 71 GAGACAGACTAGCTGATTTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGGAGG 130
 DB 1998 GAGACAGACTAGCTGATTTCTTCTAGGCGCACTAAGATCCCTAAGCTTGGGAGG 1939
 QY 131 TGACCACTGCTTAAACACGGGGCTTGCACTTACCTCAACCTGACCAATC---- 186
 DB 1938 TGACTGATCCTTAAACATGGGGCTTGCACTTACCTCAACCTGACCAATCAGGT 1879
 QY 187 -----AGAGAGCTCACTAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCAA 241
 DB 1878 ATTAAGAGAGCTCACTAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCAA 1819
 QY 242 TCATCTATTGCTGAGACAGACAGAGAGGACAAATCGGATATATAACCCAGGCAAT 301
 DB 1818 TCATCTATTGCTGAGACAGAGAGAGGACAAATCGGATATATAACCCAGGCAAT 1759
 QY 302 CGAGCTGGCAACAG 361
 DB 1758 TGAGCTGGCAACAG 1704
 QY 362 GCTATTCTACTTATTAATCTTGCACCTGCA 393
 DB 1703 -----TTTCACTCTACTTAAATCTTCAATTTGCA 1676

RESULT 13

US-09-864-761-14951/C
 Sequence 14951: Application US/09864761
 Patent No. US2002048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aeomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 PRIOR FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263.6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO: 14951
 LENGTH: 569
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AP000233.1
 OTHER INFORMATION: EXPRESSED IN PETIL LIVER, SIGNAL = 9.3
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
 US-09-864-761-14951

Query Match 40.9%; Score 204.6; DB 10; Length 569;
 Best Local Similarity 79.4%; Pred. No. 1.6e-61;
 Matches 336; Conservative 0; Mismatches 64; Indels 23; Gaps 7;

QY 66 ACTGAG 125
 DB 415 ACCGAG 356
 QY 126 G-AAGTGACCAAGCTCCACCTTTAAACACGGGGCTTGCACTTAGCTCACAGCTGACCA 184
 DB 355 GAAAGTGACCAAGCTCCACCTTTAAACACGGAGCGTGTACTAGCTCGATCGACCA 296
 QY 185 TC-----AGAGAGCTACTTAAATGCTAATTAGGC-AAAGACAGAGAGTAAAGAA 234
 DB 295 TCAGGTAGTAAAGAGGGGTTTACTGAAATACAAATTAGGCTAAAGAGAGAGTAAAGAAA 236
 QY 235 TAG-CCAATCATCTATTGCTGAGAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 292
 DB 235 TAGCAATCATCTATTGCTGAGAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 176
 QY 293 CCAAGCATTCGAGCTGCAACAG 352
 DB 175 CCAAGCATTCGAGCTGCAACAG 116
 QY 353 TGTTCATGCTATTCTACTCTTAAATCTTGAATGCACTG--CACTCTTGTGTCATGTT 410
 DB 115 TGTGT-----TTTCACTCTGTTAAATCTTGAATGCACTGTCGTCAGTGT 64
 QY 411 TCTTACGCTGAGCTGAGCTTTTGTCTACAGCTCCACCACTGCTGTTTGCACCAAGCA 470
 DB 63 TGTTCGAGCTGAGCTGAGCTTTTGTCTACAGCTCCACCACTGCTGTTTGCACCAAGCA 4
 QY 471 GAC 473
 DB 3 GAC 1

RESULT 14

US-10-040-916-50
 Sequence 50: Application US/10040916
 Patent No. US20020146769A1
 GENERAL INFORMATION:
 APPLICANT: Jacobs, Kenneth
 APPLICANT: McCoy, John
 APPLICANT: Lavallee, Edward

Racie, Lisa
Merberg, David
Treacy, Maurice
Evans, Cheryl
Spaulding, Wilki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,916
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,029
FILING DATE: 07-FEB-1997
APPLICATION NUMBER: 08/586,878
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-10-040-916-50
Query Match 40.0%; Score 200; DB 12; Length 279;
Best Local Similarity 88.2%; Pred. No. 4,66-60;
Matches 232; Conservative 1; Mismatches 27; Indels 3; Gaps 2;
QY 133 ACCAGCTCCACCTTTAAACACGGGCTTGCACTTACCTGACCTGACAGAGNG 192
DB 1 RCCACATCCACTTTAAACACGGGNTTCCAAANAGATMACACTTGACCATCAGAGNG 60
QY 193 CTCCTAAATGCTAATTAGGCAAGAGAGGTTAAATAATGCCATCTATTTC 252
DB 61 NTCANTAAATGATATTATTTGGCAAAAACGAGAGTTAAATAATGACCATCTATTTC 120
QY 233 CTGAAGACACAGAGAGAGGAGCAACAATGGGATATAAACCAGCATTCGAGCTGGCAA 312
DB 121 CTGAAGACACAGAGAGGAGCAACAATGGGATATAAACCAGATTTCAGCCGCGCAA 180
QY 313 CAGGAGCCCCCTTTGGGCTCTCCCTTTGATGGAGC--TGTTTCATCTATTTC 370
DB 181 CGGCA-ACCCCTTTGGGCTCTCCCTTTGATGGAGCCTTTGTTTCATCTATTTC 239
QY 371 CTCTATTAAATCTTGCAACTGCA 393
DB 240 NTMTATTAAATTTGCAACTGCA 262

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 20462
LENGTH: 246
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010951.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: AB026898.1, EVALUE 4.00e-57
OTHER INFORMATION: EST HUMAN HIT: A1492055.1, EVALUE 3.00e-49
OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALUE 7.30e-01
US-09-864-761-20462
Query Match 25.2%; Score 126; DB 10; Length 246;
Best Local Similarity 85.6%; Pred. No. 4,56-34;
Matches 178; Conservative 0; Mismatches 20; Indels 10; Gaps 3;

```

Qy 246 CTATTGCTGAGACGACAGAGAGGACAACAATCGGATATATAACCCAGCATTCGAG 305
    |||||
Db 246 CTGTTGCTGAGACACAGCGGAGAGGACAAATATATACAGATATATAACCCAGCATTCGAG 187
    |||||
Qy 306 CTGCAACAGCAGCCCTTTGGGTCCTTCCCTTTGTATGGAGCTGTTTCATGCTA 365
    |||||
Db 186 CTGCAACGGTA-ACCCCCTTTGGGTCCCTTGTATGGAGCTCTAT----- 136
    |||||
Qy 366 TTTCACTCTATTAATCTTGCAACTGCACCTCTTGATGCCATGTTCTTAAGGCTCGAGC 425
    |||||
Db 135 CTTCACTCTATTAATCTTGCAACTGTACTTCTTGATGCCATGTTCTTAAGGCTCGAGC 77
    |||||
Qy 426 TGAAGTTTGCTCACCGTCACCACTGC 453
    |||||
Db 76 TGAAGTTTGCTCACCGTCACCACTGC 49
    |||||

```

Search completed: April 17, 2003, 07:46:40
 Job time : 301.169 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:30:25 ; Search time 1266.31 seconds
(Without alignments)
6394.748 Million cell updates/sec

Title: US-09-719-554-3_COPY_1_500

Perfect score: 500

Sequence: 1 cccctggggggggctccctt.....tgactccatccctcgat 500

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estsda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	336	67.2	727	17	AG029908 Pan trogl
2	322.2	64.4	683	17	AG102951 Pan trogl
3	285.4	57.1	440	12	BE732673 BE732673
4	283.8	56.8	415	9	AI128496 AI128496
5	283.2	56.6	436	9	AI128526 AI128526
6	282.2	56.4	494	9	AA781423 AA781423

C	7	280	56.0	342	9	AA660368	AA660368 aj59c05.s
C	8	277.2	55.4	921	17	BH149565	BH149565 ENT0948TR
C	9	273	54.6	422	14	NS3177	NS3177 yv6h11.s1
C	10	271.8	54.4	388	14	H01325	H01325 y19e01.s1
C	11	271.2	54.2	385	14	NS5091	NS5091 yv43e03.s1
C	12	270.8	54.2	342	14	T47345	T47345 yb10h02.s1
C	13	270.6	54.1	771	13	BI087886	BI087886 602852690
C	14	267.4	53.5	522	10	AW971553	AW971553 EST183642
C	15	265	53.0	363	14	T69704	T69704 yd13a03.s
C	16	263.6	52.7	328	9	AA729556	AA729556 nx58c05.s
C	17	261.4	52.3	410	9	AA250958	AA250958 z807d10.s
C	18	261	52.2	443	9	AA837267	AA837267 o426d10.s
C	19	260.4	52.1	681	10	AV722664	AV722664 AV722664
C	20	259.6	51.9	425	9	A1570707	A1570707 tm79g09.x
C	21	259.6	51.9	433	9	A1379210	A1379210 cd01g11.x
C	22	259.6	51.9	490	9	A1598135	A1598135 tn14a10.x
C	23	258.8	51.8	681	10	AV721910	AV721910 AV721910
C	24	258.2	51.6	424	14	R27412	R27412 yb46d11.s1
C	25	257.6	51.5	458	14	R76086	R76086 y171b03.s1
C	26	256.2	51.2	404	14	R27389	R27389 yb46a09.s1
C	27	253.4	50.7	893	14	BQ437925	BQ437925 AGENCOURT
C	28	253.2	50.6	446	9	A1393478	A1393478 tg45g04.x
C	29	250.8	50.2	611	17	AQ381711	AQ381711 RPCI11-16
C	30	246.6	49.3	332	9	A1797629	A1797629 w82a08.x
C	31	245.4	49.1	701	17	AG126669	AG126669 Pan trogl
C	32	243.8	48.8	438	14	R77278	R77278 y175d06.s1
C	33	243.6	48.7	722	17	AG104643	AG104643 Pan trogl
C	34	243.4	48.7	396	9	AA814939	AA814939 oc07d02.s
C	35	239	47.8	722	17	AG049481	AG049481 Pan trogl
C	36	238.6	47.7	326	14	D29167	D29167 HUNNK203 Hu
C	37	238.6	47.7	463	14	R68685	R68685 y114g06.s1
C	38	237.6	47.5	447	12	BF919416	BF919416 QV0-NT015
C	39	236.8	47.4	683	17	AG134524	AG134524 Pan trogl
C	40	234.8	47.0	330	9	AA774109	AA774109 ac36d03.s
C	41	233.6	46.7	470	9	A1791155	A1791155 ab52f07.x
C	42	233.6	46.7	440	12	BF919425	BF919425 QV0-NT015
C	43	232.6	46.5	440	9	A1003607	A1003607 zE99a12.s
C	44	232.6	46.5	471	9	AA709471	AA709471 zE91h06.s
C	45	232.4	46.5	757	17	AG121490	AG121490 Pan trogl

ALIGNMENTS

RESULT 1	AG029908	727 bp	DNA	linear	GSS 01-NOV-2001
LOCUS	Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.				
DEFINITION	AG029908				
ACCESSION	AG029908.1	GI:16556781			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
Pan troglodytes male lymphoblast DNA, clone: PTB Chimpanzee Male					
BAC library clone: PTB-002C04.R.					
Pan troglodytes					
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.					
REFERENCE					
1	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,				
Totoki, Y., Watanabe, H. and Sakaki, Y.					
BAC end sequences of library PTB					
Unpublished					
2	(bases 1 to 727)				
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,					
Totoki, Y., Watanabe, H. and Sakaki, Y.					
Direct Submission					
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical					
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),					
1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan					
(E-mail: chimboes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,					
Tel: 81-45-503-9111, Fax: 81-45-503-9170)					
Clones are derived from the chimpanzee BAC library PTB This BAC end					
was generated during the R&D process and may have higher chance of					
clone tracking errors.					
TITLE	JOURNAL				
COMMENT					

PRIMERS

Sequencing: M13Rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

FEATURES

source

1..727

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-002C04.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC library"

BASE COUNT 170 a 215 c 175 g 166 t 1 others

ORIGIN

Query Match 67.2%; Score 336; EB 17; Length 727;

Best Local Similarity 87.2%; Pred. No. 3,6e-99;

Matches 380; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 61 TTGCACTGAGAGACGAGCTAGCTGATTTCTGAGCCGAGTAAAGATCCCTAAGCCTA 120

Db 270 TTCTCATGAGAGACGAGCTAGCTGATTTCTGAGCCGAGTAAAGATCCCTAAGCCTA 329

QY 121 GCTGGGAGGTGACGACGCTTAAACAGCGGGTTTGCACTTACCTCACAACCTGA 180

Db 330 GCTGGGAGGTGACGACGCTTAAACAGCGGGTTTGCACTTACCTCACAACCTGA 389

QY 181 CCAATCAGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCA 240

Db 390 CCTATCGGAGAGCTCACTAAATGCTAATTAGGCAAAACAGAGAGTAAAGAAATAGCCA 449

QY 241 ATCATCTATGCTGAGAGCAAGAGAGGAGCAACATCGGAGTATTAACCCAGGCAAT 300

Db 450 ATCATCTATGCTGAGAGCAAGAGAGGAGCAACATCGGAGTATTAACCCAGGCAAT 509

QY 301 TCGAGCTGGCAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTTCA 360

Db 510 TCAAGCCAGCAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTTCA 569

QY 361 TCGTATTTGACTTATTAATCTTGCACTGCACTCTTGTATGAGAGCTGTTTCA 420

Db 570 CTCTATTTGACTTATTAATCTTGCACTGCACTCTTGTATGAGAGCTGTTTCA 629

QY 421 CGAGTGAGCTTTTGTCTGACCGTCCACCACTGCTT-TGCCACCAACCGCAACCTGCCG 479

Db 630 CGAGTGAGCTTTTGTCTGACCGTCCACCACTGCTT-TGCCACCAACCGCAACCTGCCG 689

QY 480 CTGACTCCCATCCCTC 495

Db 690 TTGACTTCATCCCTC 705

RESULT 2

AG102951

LOCUS AG102951 683 bp DVA linear GSS 03-NOV-2001

DEFINITION Pan troglodytes DNA, clone: PTB-106G16.F, genomic survey sequence.

ACCESSION AG102951.1 GI:16723468

KEYWORDS

Pan troglodytes male lymphoblast DNA, clone: PTB Chimpanzee Male

BAC library clone: PTB-106G16.F.

SOURCE

ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

Tokoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of library PTB

Unpublished

2 (bases 1 to 683)

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,

AUTHORS

TITLE

Direct Submission

Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical

and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

1-7-22 Suehiro-chou, Tsukuba, Ibaraki, Japan

(E-mail: chimpbes@gsr.riken.go.jp, URL: http://hsp.gsc.riken.go.jp/)

Tel: 81-45-503-9111, Fax: 81-45-503-9170

Clones are derived from the chimpanzee BAC library PTB This BAC end

was generated during the R&D process and may have higher chance of

clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1..683

/organism="Pan troglodytes"

/db_xref="taxon:9598"

/clone="PTB-106G16.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC library"

BASE COUNT 162 a 186 c 167 g 167 t 1 others

ORIGIN

Query Match 64.4%; Score 322.2; DB 17; Length 683;

Best Local Similarity 87.4%; Pred. No. 1.2e-94;

Matches 387; Conservative 0; Mismatches 53; Indels 3; Gaps 3;

QY 59 TCTTGAAGTGAAGAGACGAGCTAGCTGATTTCTTGGCCGAGTAAAGATCCCTAAGCC 118

Db 101 TCTTGAAGTGAAGAGACGAGCTAGCTGATTTCTTGGCCGAGTAAAGATCCCTAAGCC 160

QY 119 TAGCTGGAGAGTGAACAGCTTAAACAGCGGGCTTGCACTTACCTCACAACCT 178

Db 161 TAGCTGGAGAGTGAACAGCTTAAACAGCGGGCTTGCACTTACCTCACAACCT 220

QY 179 GACCAATGAGAGAGTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAGAAATAGC 238

Db 221 AACCAATGAGAGAGTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAGAAATAGC 279

QY 239 CAATCATATATGCTGAGAGACAGAGAGGAGCAACATCGGAGATATAACCCAGGC 298

Db 280 CAATCATATATGCTGAGAGACAGAGAGGAGCAACATCGGAGATATAACCCAGGC 339

QY 299 ATTGAGCTGGGAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTT 358

Db 340 ATTGAGCTGGGAACAGAGCCCTTTGGGCTCCCTTGTATGAGAGCTGTTT 398

QY 359 CATGCTATTTGACTTATTAATCTTGCACTGCACTCTTGTATGAGAGCTGTTT 418

Db 399 CACTTATTTGACTTATTAATCTTGCACTGCACTCTTGTATGAGAGCTGTTT 458

QY 419 CTCGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTG-TTTCACCAACCGCAACCTGC 477

Db 459 CTCGAGCTGAGCTTTTGTCTACCGTCCACCACTGCTG-TTTCACCAACCGCAACCTGC 518

QY 478 CGCTGACTCCCATCCCTCTGAT 500

Db 519 CGCTGACTCCCATCCCTCTGAT 541

RESULT 3

BE732673

LOCUS BE732673 440 bp mRNA linear EST 15-SEP-2000

DEFINITION 601571305F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3925728 5',

mRNA sequence.

ACCESSION BE732673

VERSION BE732673.1 GI:10146665

KEYWORDS

human.

ACCESSION A1128526
VERSION A1128526.1 GI:3597040
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 720 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amerisham
High quality sequence stop: 428.
Location/Qualifiers
1. 436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1714147"
/clone_lib="Soares placenta, 8to9weeks 2NDHP8to9W"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTGGAGCGGCGCGATTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 95 a 96 c 106 g 139 t
ORIGIN
Query Match 56.6%; Score 283.2; DB 9; Length 436;
Best Local Similarity 95.2%; Pred. No. 7e-02;
Matches 314; Conservative 0; Mismatches 13; Indels 3; Gaps 2;
QY 66 ACTGAGAGCAGAGCTAGCTGATTTCTAGGCCGACCTAAGATCCCTAAGCTAGCTGG 125
DB 351 ACTGAAAGCAGAGCTAGCTGATTTCTAGGCCGACCTAAGATCCCTAAGCTAGCTGG 292
QY 126 GAAGGTGACACGCTCCACTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAAT 185
DB 291 GAAGGTGACACATCCACTTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAAT 232
QY 186 CAGAGAGCTCAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 245
DB 231 CAGAGAGCTCAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 172
QY 246 CTAATGCCCTGAGAGCAGACAGAGGAGCAACATGCGATTAATTAACCCAGCATTCGAG 305
DB 171 CTAATGCCCTGAGAGCAGACAGAGGAGCAACATGCGATTAATTAACCCAGCATTCGAG 112
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCTTTGATGGAG--CTGTTTTCATGC 363
DB 111 CCGGCAACGGCA-ACCCCTTTGGGTCCCTTCCTTTGATGGAGCTGTGTTTCATGC 53
QY 364 TATTCTACTCTAATTAATCTTGCACTGCA 393
DB 52 TATTCTACTCTAATTAATCTTGCACTGCA 23
RESULT 6
AA781423/c AA781423 494 bp mRNA linear EST 31-DEC-1998
LOCUS
DEFINITION aj26c03.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'

similar to contains PTR7.ct1 PTR7 repetitive element ;, mRNA sequence.
ACCESSION AA781423
VERSION AA781423.1 GI:2840754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bdtp/image/image.html
Insert Length: 1645 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amerisham
High quality sequence stop: 475.
Location/Qualifiers
1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1391428"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCATCTGAAGTGGAGCGGCGCGATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 108 a 111 c 133 g 141 t
ORIGIN
Query Match 56.4%; Score 282.2; DB 9; Length 494;
Best Local Similarity 94.8%; Pred. No. 1.6e-81;
Matches 313; Conservative 0; Mismatches 14; Indels 3; Gaps 2;
QY 66 ACTGAGAGCAGAGCTAGCTGATTTCTAGGCCGACCTAAGATCCCTAAGCTAGCTGG 125
DB 334 ACTGAGAGCAGAGCTAGCTGATTTCTAGGCCGACCTAAGATCCCTAAGCTAGCTGG 275
QY 126 GAAGGTGACACGCTCCACTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAAT 185
DB 274 GAAGGTGACACATCCACTTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAAT 215
QY 214 CAGAGAGCTCAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 245
DB 186 CAGAGAGCTCAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCAT 155
QY 246 CTAATGCCCTGAGAGCAGACAGAGGAGCAACATGCGGATTAATTAACCCAGCATTCGAG 305
DB 154 CTAATGCCCTGAGAGCAGACAGAGGAGCAACATGCGGATTAATTAACCCAGCATTCGAG 95
QY 306 CTGGCAACAGACAGCCCTTTGGGTCCCTTCCTTTGATGGAG--CTGTTTTCATGC 363
DB 94 CCGGCAACGGCA-ACCCCTTTGGGTCCCTTCCTTTGATGGAGCTGTGTTTCATGC 36
QY 364 TATTCTACTCTAATTAATCTTGCACTGCA 393

Db 35 TATTCACTCTATTAAATCTTGCAACTGCA 6

RESULT 7
AA860368/c
LOCUS 342 bp mRNA linear EST 31-DEC-1998
DEFINITION aj59c05.s1 Soares testis NHT Homo sapiens CDNA clone IMAGE:1394600
3' similar to contains PTR7.c1 PTR5 repetitive element ;, mRNA
sequence.

ACCESSION AA860368
VERSION AA860368
KEYWORDS EST.
SOURCE AA860368.1 GI:2954363
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 342)
NCI-CCAP htcp://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
www.bio.linn.gov/bdtp/image/image.html
Insert Length: 1305 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 319.
Location/Qualifiers

FEATURES
source
1..342
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394600"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pTR73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGACATGCGAGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR73 vector. Library
went through one round of normalization to Cos, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 75 a 78 c 82 g 107 t

ORIGIN

Query Match 56.0%; Score 280; DB 9; Length 342;
Best Local Similarity 94.5%; Pred. No. 7.1e-81;
Matches 312; Conservative 0; Mismatches 15; Indels 3; Gaps 2;

Db 66 ACTGAGACGAGACTAGCTGATTTCTAGCCGCACTAAAGATTCCTAAGCTTACTGG 125
338 ACTGAGACGAGACTAGCTGATTTCTAGCCGCACTAAAGATTCCTAAGCTTACTGG 279

Db 126 GAAGTGACGACGCTGACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
278 GAAGTGACGACCTGACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 219

Db 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGACCAATCAT 245
218 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATGACCAATCAT 159

Db 246 CTATTGCTGAGACGACGAGGAGCAACATCGGAGATTAAACCCAGCATTCGAG 305
158 CTATTGCTGAGACGACGAGGAGCAATGATCGGAGATTAAACCCAGCATTCGGA 99

Qy 306 CTGGCAACAGCAGCCCCCTTGGGCTCCCTTGTATGGAG--CTGTTTCATGC 363
Db 98 GCGGCAACGGCA-AACCCCTTTGGGCTCCCTTGTATGGAGCTGTGTTTATGC 40

Qy 364 TATTCACTCTATTAAATCTTGCAACTGCA 393
Db 39 TATTCACTCTATTAAATCTTGCAACTGCA 10

RESULT 8
BH149565/c
LOCUS 921 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTQ648TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
BH149565
ACCESSION BH149565.1 GI:15310303
VERSION GSS.
KEYWORDS Entamoeba histolytica.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 921)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HML:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HML:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 40
High quality sequence stop: 567.
Location/Qualifiers

FEATURES
source
1..921
/organism="Entamoeba histolytica"
/strain="HML:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/note="Vector: pHOSt, Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barell, Oxford University Press, 1999). "

BASE COUNT 242 a 216 c 226 g 237 t

ORIGIN

Query Match 55.4%; Score 277.2; DB 17; Length 921;
Best Local Similarity 84.5%; Pred. No. 9.1e-80;
Matches 377; Conservative 0; Mismatches 48; Indels 21; Gaps 5;

Db 66 ACTGAGACGAGACTAGCTGATTTCTAGCCGCACTAAAGATTCCTAAGCTTACTGG 125
651 AGTGAAGACAGGACTAGCTGATATCTTAGGCTTAAGAAATCTTAAGCTTACTGG 592

Db 126 GAAGTGACGACGCTGACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 185
591 GAAGTGACGACCTGACCTTTAAACAGGGGCTTGCAACTAGCTACACCTGACCAAT 532

OY	186	C-----AGAGGCGCTCATTAAATGGTATTTCAGGCAAAACAGAGGGTTAAAGAATA	236
Db	531	CAGTAGTAAGGAAGGCTACTTAAAACGTAAATTAGGCAGAAACAGAGAGTTAAGAAATTA	472
OY	237	GCCATTCATCTATTTCCTGTAGAGCACAGACAGAGGAGACACAACATCGGGATTAATTAACCCAG	296
Db	471	GCCATTCATCTATTTCCTGTAGAGCACACCTGGAGAGACATATAT-AGGATTAATTAACCCAG	413
OY	297	GCATTGCGAGCTGGCAACAGACAGCCCCCTTTGGGTCCTTCCCTTTGTAATGGAGACTGTT	356
Db	412	GCATTCAAACCGACAGAAAGGCG-TACCCTCTTTGAGTCCCCTCCCTTTGTAATGGAGACTGTG	354
OY	357	TTCATGCAATTTACCTCATTAATCTTGCACTG--CACTCTTCGGGCCATGTTTTCTT	414
Db	353	T-----TTTACCTCATTAATCTTGCACTG-CACCTCTTCTTGCACTGTTTTGTT	302
OY	415	ACGGCTCGAGCTGAGCTTTTGCTCAACCGTCACCA-TGCTGTTTTCACCAACCGAGACC	474
Db	301	ACGACTCGAGCTGAGCTTTTCACTTCGCGTCCACCAATGCTGTTTTCCGCGCATCGAGACC	242
OY	475	TGCGCGTGACTCCCATCCCTCTTGAT	500
Db	241	CGCCGCTGACTTCCACCCCTCCAGAT	216
RESULT_9	N53177/c	N53177	422 bp mRNA linear EST 28-JAN-1997
LOCUS		yv6h11.b1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone	
DEFINITION		IMAGE:246789.3 similar to contains PTRJ.t2 PVR7 repetitive element	
ACCESSION		N53177	
VERSION		N53177.1	GI:1194343
KEYWORDS		EST.	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Cran.tra; Vertebrata; Euteleostomi;	
REFERENCE		Mammalia; Euteria; Primates; Carnivora; Homnidae; Homo.	
AUTHORS		1 (bases 1 to 422) Hillier,L., Lennon,G., Becker,M., Iacono,M.F., Chiapelli,B., Chisoso,S., Dietrich,N., Dubuque,T., Favallo,A., Gish,W., Hawkins ,M., Hillman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore, ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996)	
TITLE		Genome Res. 6 (9), 807-828 (1996)	
JOURNAL		97044478	
MEDLINE		Contact: Wilson RK	
COMMENT		Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 938 Std Error: 0.00 Seq primer: ml3 -40 forward High quality sequence strop: 301.	
FEATURES		location/Qualifiers	
source		1..422	
		/organism="Homo sapiens"	
		/db_xref="GDB:3796035"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:246789"	
		/clone_lib="Soares fetal liver spleen INFLS"	
		/sex="male"	
		/dev_stage="20 week-post conception fetus"	
		/lab_host="DH10B (ampicillin resistant)"	
		/note="Organ: Liver and Spleen; Vector: pTR73D (Pharmacia)	
		with a modified polylinker; Site 1: Pac I; Site 2: Eco RI	
		1st strand cDNA was primed with a Pac I - oligo(dT) primer	
		[5'] AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'1,	

BASE COUNT		95 a		95 c		111 g		120 t		1 others	
ORIGIN											
Query Match 54.6%; Score 273; DB 14; Length 422;											
Best Local Similarity 93.3%; Pred. No. 1,66-76;											
Matches 308; Conservative 0; Mismatches 16; Indels 6; Gaps 2;											
QY	66	ACTGAGAGACAGACTGAGATTTCCTTAGCCGACATAAGATCCCTAAGCTTAGCTTG	125								
DB	327	ACTGAGAGACAGACTGAGATTTCCTTAGCCGACATAAGATCCCTAAGCTTAGCTTG	268								
QY	126	GAAAGTGACCAAGTCCACCTTTAAACAGCGGGGCTTGCACTTAGCTCACACCTGACCAAT	185								
DB	267	GAAAGTGACCAAGTCCACCTTTAAACAGCGGGGCTTGCACTTAGCTCACACCTGACCAAT	208								
QY	186	CAGAGAGCTCATTAAATGCTTAATAGCCAAAGACAGAGGTAAAGAAATAGCCAAATAT	245								
DB	207	CAGAGAGCTCATTAAATGCTTAATAGCCAAAGACAGAGGTAAAGAAATAGCCAAATAT	148								
QY	246	CTATTGCTTGAGAGACAGACAGAGGAGCAACAATCGGATATATAACCCAGGATTTTCAG	305								
DB	147	CTATTGCTTGAGAGACAGACAGAGGAGCAACAATGATCGGATATATAACCCAGGATTTTCAG	88								
QY	306	CTGGCAGACAGCAGCCCCCTTTGGGTCCCTTCCTTTGTATGGAG--CTGTTTTCATGC	363								
DB	87	CGG----CAAGGGCAACCCCTTTGGTCCCTTCCTTTGTATGGAGCTGTGTTTCATGC	32								
QY	364	TATTTCACCTCTATTAAATCTTGCAACTGCA	393								
DB	31	TATTTCACCTATTAAATCTTGCAACTGCA	2								
RESULT 10											
LOCUS H01325 388 bp mRNA linear EST 19-JUN-1995											
DEFINITION y199e01.s1 Soares placenta N02HP Homo sapiens cDNA clone											
IMAGE:147384 3', mRNA sequence.											
ACCESSION H01325											
VERSION H01325.1 GI:864258											
KEYWORDS EST.											
SOURCE human.											
ORGANISM Homo sapiens											
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.											
REFERENCE 1 (bases 1 to 388)											
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maitra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterson,R., Williamson,A., Wohlmann,P. and Wilson,R.											
The Washu-Merck EST Project											
TITLE Unpublished (1995)											
JOURNAL											
COMMENT											
Contact: Wilson RK											
Washington University School of Medicine											
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108											
Tel: 314 286 1800											
Fax: 314 286 1810											
Email: est@watson.wustl.edu											
Insert Size: 790											
High quality sequence stops: 146											
Source: IMAGE Consortium, LNLN											
This clone is available royalty-free through LNLN; contact the											
IMAGE Consortium (info@image.lnl.nyu.edu) for further information.											
Insert Length: 790 Std Error: 0.00											
Seq primer: Promega -21ml3											
High quality sequence stop: 346.											
Location/Qualifiers											
1..388											
/Organism="Homo sapiens"											

/db_xref="GDB:559031"
 /db_xref="taxon:9606"
 /clone="IMAGE:147384"
 /clone_lib="Soares placenta Nb2HP"
 /sex="Female"
 /dev_stage="Placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', ACTGGAAGAATTGCGGCGCAGAGATTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 90 a 90 c 96 g 110 t 2 others
 ORIGIN

Query Match 54.4%; Score 271.8; DB 14; Length 388;
 Best Local Similarity 93.8%; Pred. No. 3.8e-78;
 Matches 305; Conservative 0; Mismatches 17; Indels 3; Gaps 2;

QY 66 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCAGCTAAGAAATCCTAAGCCTTAGCTGG 125
 Db 329 ACTGAGAGACGAGACTGCTGATTTCTTAGCTGAGCTAAGAAATCCTAAGCCTTAGCTGG 270
 QY 126 GAAGGTGACCGACGTCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGCTGACCAAT 185
 Db 269 GAAGGTGACCGACATCCACTTTAAACAGGGGGCTTGCACTTAGCTCAGCTGACCAAT 210
 QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATGCCAATCAT 245
 Db 209 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCAT 150
 QY 246 CTATTGCTTAGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 305
 Db 149 CTATTGCTTAGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 90
 QY 306 CTGGCAACAGAGCGCCCTTTGGTGCCCTTCCCTTTGATAGGAG--CTGTTTTCATGC 363
 Db 89 GCGGCAACGGCA-AGCCCTTTGGTGCCCTTCCCTTTGATAGGAGCTCTGTTTTCATGC 31
 QY 364 TATTCACTCTATTAAATCTTGCA 388
 Db 30 TATTCACTCTATTAAATCTTGCA 6

RESULT 11 385 bp mRNA linear EST 28-JAN-1997
 N55091/c
 LOCUS y743803.g1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 DEFINITION IMAGE:245500 3', mRNA sequence.

ACCESSION N55091
 VERSION N55091.1 GI:1197970
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 385)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., Dubucq, T., Pavello, A., Gibb, W., Hawkins, M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Merra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

TITLE JOURNAL MEDLINE
 CONTACT: Wilson RK
 WASHINGTON UNIVERSITY SCHOOL OF MEDICINE
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 907 Std Error: 0.00
 Seq primer: m13 -40 forward
 High quality sequence stop: 276.
 location/Qualifiers
 1. 385

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="GDB:3794746"
 /db_xref="taxon:9606"
 /clone="IMAGE:245500"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', ACTGGAAGAATTATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 88 a 88 c 102 g 106 t 1 others
 ORIGIN

Query Match 54.2%; Score 271.2; DB 14; Length 385;
 Best Local Similarity 93.0%; Pred. No. 5.9e-78;
 Matches 306; Conservative 0; Mismatches 19; Indels 4; Gaps 2;

QY 66 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCAGCTAAGAAATCCTAAGCCTTAGCTGG 125
 Db 327 ACTGAGAGACGAGACTGCTGATTTCTTAGCGCTGAGCTAAGAAATCCTAAGCCTTAGCTGG 268
 QY 126 GAAGGTGACCGACGTCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGCTGACCAAT 185
 Db 267 GAAGGTGACCGACATCCACTTTAAACAGGGGGCTTGCACTTAGCTCAGCTGACCAAT 208
 QY 186 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCAT 245
 Db 207 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCAT 148
 QY 246 CTATTGCTTAGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 305
 Db 147 CTATTGCTTAGAGACAGCAGAGAGGACAAATCGGATATTAACCCAGGCAATTGAG 88
 QY 306 CTGGCAACAGAGCGCCCTTTGGTGCCCTTCCCTTTGATAGGAG--CTGTTTTCATGC 363
 Db 87 C-GGCAACGGCAACCCCTTTGGTGCCCTTCCCTTTGATAGGAGCTCTGTTTTCATGC 30
 QY 364 TATTCACTCTATTAAATCTTGCAATGC 392
 Db 29 TATTCACTCTATTAAATCTTGCAATGC 1

RESULT 12 342 bp mRNA linear EST 01-FEB-1995
 LOCUS T47345
 DEFINITION yb10h02.g1 Stragagene placenta (#937225) Homo sapiens cDNA clone
 IMAGE:70803 3', mRNA sequence.

ACCESSION T47345
 VERSION T47345.1 GI:649326
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 342)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

SUMMARIES

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Run on: April 19, 2003, 10:07:23 ; Search time 1703.47 Seconds

8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Sequence: 1 gtctcggccaacctcccaa.....ctggaatcatgcaggtgtcc 501

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Maximum DB seg length: 2000000000
Minimum DB seg length: 0
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%

Listing first 45 summaries

Database

1:	gb_bml:*
2:	gb_htg:*
3:	gb_in:*
4:	gb_cm:*
5:	gb_ov:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gd_ro:*
11:	gd_stc:*
12:	gb_sun:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hun:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_ox:*
22:	em_ov:*
23:	em_pac:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_sta:*
28:	em_un:*
29:	em_vl:*
30:	em_htg_hum:*
31:	em_htg_iny:*
32:	em_htg_other:*
33:	em_htg_mus:*
34:	em_htg_pln:*
35:	em_htg_rod:*
36:	em_htg_mam:*
37:	em_htg_vtc:*
38:	em_sy:*
39:	em_htg_hum:*
40:	em_htg_mus:*
41:	em_htg_other:

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	501	100.0	10499	6	AX007980	AX007980 Sequence
2	498	99.4	56093	6	AX329572	AX329572 Sequence
3	498	99.4	56093	6	HSAC000664	AC000664 Human BAC
4	498	99.4	149194	9	AC007566	AC007566 Homo sapi
5	450	89.8	711	6	AX007998	AX007998 Sequence
6	434	86.6	174019	9	AP001538	AP001538 Homo sapi
7	434	86.6	340000	9	AP001674	AP001674 Homo sapi
8	431.2	86.1	163803	9	AC093551	AC093551 Homo sapi
9	426	85.0	839	9	AF127226	AF127226 Homo sapi
10	426	85.0	1329	6	AX001030	AX001030 Sequence
11	425.6	85.0	783	6	AX000970	AX000970 Sequence
12	415.6	83.0	251124	9	HUAE000660	ABE00660 Homo sapi
13	409.6	81.8	187321	2	AC092510	AC092510 Homo cyn
14	409	81.6	83412	9	AC092843	AC092843 Homo sapi
15	407.2	81.3	140756	9	AL139038	AL139038 Human DNA
16	407.2	81.3	143499	2	AL1607153	AL1607153 Homo sapi
17	400.6	80.0	169029	9	CNS06C7R	AL1390800 Human chr
18	400.2	79.9	110000	2	AL353584_0	AL353584 Homo sapi
19	400.2	79.9	170746	9	AL392173	AL392173 Human DNA
20	398.8	79.6	151863	9	AC010888	AC010888 Homo sapi
21	396.6	79.2	88328	9	AL357874	AL357874 Human DNA
22	396.4	79.1	157711	9	AC055738	AC055738 Homo sapi
23	396.4	79.1	207945	9	CNS01DSS	AL121841 Human chr
24	396.4	79.1	318442	2	AC021652	AC021652 Homo sapi
25	396.2	79.1	771	6	AX007997	AX007997 Sequence
26	392.8	78.4	114455	9	AC093742	AC093742 Homo sapi
27	391.4	78.1	165059	9	AC106856	AC106856 Homo sapi
28	390	77.8	137492	2	AL161721	AL161721 Homo sapi
29	389.8	77.8	181753	9	AC124075	AC124075 Homo sapi
30	389.6	77.8	143590	9	HS834075	AL035706 Human DNA
31	388	77.4	119406	2	AC103595	AC103595 Homo sapi
32	387.4	77.3	76169	9	AC003014	AC003014 Human PAC
33	386.8	77.2	180635	9	AL360169	AL360169 Human DNA
34	386	77.0	145785	2	AC016173	AC016173 Homo sapi
35	385.8	76.9	161049	2	AL1591842	AL1591842 Homo sapi
36	385.4	76.9	111079	9	HS4193C19	AL035407 Human DNA
37	385.2	76.9	156177	9	AC093896	AC093896 Homo sapi
38	384.8	76.8	42216	9	AC092899	AC092899 Homo sapi
39	384.8	76.8	170414	2	AC026019	AC026019 Homo sapi
40	384.6	76.8	119481	9	AL136233	AL136233 Human DNA
41	384.4	76.7	91097	9	AP001600	AP001600 Homo sapi
42	384.4	76.7	118241	9	AP001599	AP001599 Homo sapi
43	384.4	76.7	185018	9	AC110614	AC110614 Homo sapi
44	384.4	76.7	340000	9	AP001697	AP001697 Homo sapi
45	383	76.4	161850	9	AC018680	AC018680 Homo sapi

JOURNAL

Patent: WO 9967395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)

FEATURES

Location/Qualifiers

1..10499

/db_xref="taxon:9606"

/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 1495 t
ORIGIN

Query Match 100.0%; Score 511; DB 6; Length 10499;

Best Local Similarity 100.0%; Pred. NC. 6.2e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTCTGGCCCACTCCCAACAGCACTAGGTTTCCTTGGATGGGGAGCTAGAGAG 60
Db 9500 GTCTGGCCCACTCCCAACAGCACTAGGTTTCCTTGGATGGGGAGCTAGAGAG 9559
Oy 61 CAGCACTAGCTGATTTCTTGAAGTGAAGATCCCTAAGCTTGAAGGAGTAC 120
Db 9560 CAGCACTAGCTGATTTCTTGAAGTGAAGATCCCTAAGCTTGAAGGAGTAC 9619
Oy 121 CACATCCACCTTTAAACAGAGGGGCTTGAAGTGAAGTGAAGTGAAGTGAAGT 180
Db 9620 CACATCCACCTTTAAACAGAGGGGCTTGAAGTGAAGTGAAGTGAAGTGAAGT 9679
Oy 181 CACTAAATGCTAATTAAGCAAGAGAGGTAAGAATAGCAATCATCTATTGCTT 240
Db 9680 CACTAAATGCTAATTAAGCAAGAGAGGTAAGAATAGCAATCATCTATTGCTT 9739
Oy 241 GAGAGCAGCAGAGAGAGCAATGATCGGATATTAACCAAGCTTTCAGCCGAGCAG 300
Db 9740 GAGAGCAGCAGAGAGAGCAATGATCGGATATTAACCAAGCTTTCAGCCGAGCAG 9799
Oy 301 GGAACCCCTTTGGGTCCTCCCTTGTATGAGGAGTCTGTTTTCATGCTATTGCTC 360
Db 9800 GGAACCCCTTTGGGTCCTCCCTTGTATGAGGAGTCTGTTTTCATGCTATTGCTC 9859
Oy 361 TATTAAATCTTGAACCTGCTTCTGATGCTATGCTTGAAGGCTTGAAGGCTT 420
Db 9860 TATTAAATCTTGAACCTGCTTCTGATGCTATGCTTGAAGGCTTGAAGGCTT 9919
Oy 421 CGCTGCGCATCCACCACTGCTGTTTGCAGCAGCCGAGACCGCGCTGATCCCATGCC 480
Db 9920 CGCTGCGCATCCACCACTGCTGTTTGCAGCAGCCGAGACCGCGCTGATCCCATGCC 9979
Oy 481 TCTGATCATGCAAGGCTGCC 501
Db 9980 TCTGATCATGCAAGGCTGCC 10000

RESULT 2

AX329572 56093 bp DNA linear PAT 09-JAN-2002

LOCUS AX329572 Sequence 81 from Patent WO0194629.

DEFINITION AX329572

ACCESSION AX329572

VERSION AX329572.1 GI:18102550

KEYWORDS

.

SOURCE

.

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Hortigan, S., Soper, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 81 13-DEC-2001;

TITLE

Avalon Pharmaceuticals (US)

JOURNAL

FEATURES

1..56093
Location/Qualifiers
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match

Best Local Similarity 99.4%; Score 498; DB 6; Length 56093;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 TCGGCCAAGCTCCCAACAGCACTAGGTTTCTGTTGATGGGGAGCTAGAGAG 63
Db 37504 TCGGCCAAGCTCCCAACAGCACTAGGTTTCTGTTGATGGGGAGCTAGAGAG 37563
Oy 64 GACTAGCTGATTTCTTGAAGTGAAGTTCCTTAAGCTTGAAGTGAAGTGAAGT 123
Db 37564 GACTAGCTGATTTCTTGAAGTGAAGTTCCTTAAGCTTGAAGTGAAGTGAAGT 37623
Oy 124 ATCCACCTTTAAACAGAGGGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 183
Db 37624 ATCCACCTTTAAACAGAGGGGCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 37683
Oy 184 TAAATGCTAATTAAGCAAGAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
Db 37684 TAAATGCTAATTAAGCAAGAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 37743
Oy 244 AGCAGCAGAGAGAGAGCAATGATGGGATATTAACCAAGTCTTGAAGCCGAGCAG 303
Db 37744 AGCAGCAGAGAGAGAGCAATGATGGGATATTAACCAAGTCTTGAAGCCGAGCAG 37803
Oy 304 ACCCCCTTTGGGTCCTCCCTTGTATGAGAGTCTGTTTTCATGCTATTGCTAT 363
Db 37804 ACCCCCTTTGGGTCCTCCCTTGTATGAGAGTCTGTTTTCATGCTATTGCTAT 37863
Oy 364 TAAATCTGCAACCTGCACTTCTGTCATGCTTCTTGAAGGCTTGAAGTGAAGT 423
Db 37864 TAAATCTGCAACCTGCACTTCTGTCATGCTTCTTGAAGGCTTGAAGTGAAGT 37923
Oy 424 TCGCATTCACCACTGCTGTTTGCAGCAGCCGAGACCGCGCTGATCCCATGCC 483
Db 37924 TCGCATTCACCACTGCTGTTTGCAGCAGCCGAGACCGCGCTGATCCCATGCC 37983
Oy 484 GGATCATGCAAGGCTGCC 501
Db 37984 GGATCATGCAAGGCTGCC 38001

RESULT 3

HSAC000064 56093 bp DNA linear PRI 13-NOV-1996

LOCUS HSAC000064 Human BAC clone RG083M05 from 7q21-7q22, complete sequence.

DEFINITION AC000064

ACCESSION AC000064.1 GI:1669369

KEYWORDS

HTG.

SOURCE

ORGANISM

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 56093)
The sequence of H. sapiens BAC clone RG083M05
unpublished (1996)

JOURNAL

TITLE

2 (bases 1 to 56093)
Direct Submission
Submitted (13-NOV-1996)

REFERENCE

Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA

e-mail: sapiens@wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

```

complement(12612. 12907)
/rpt_family="ALU"
13670. .13793
/gene="WUGSC:H.RG083M05.1"
/note="match to human 5' EST H41382 (NID:917434), bases
143-266"
13794. .13877
/rpt_family="ALU"
13878. .13906
/gene="WUGSC:R.RG03M05.1"
/note="match to human 5' EST H41382 (NID:917434), bases
30-58"
13907. .14104
/rpt_family="ALU"
complement(14110. .14137)
/rpt_family="L1"
complement(15618. .15907)
/rpt_family="ALU"
17227. .17522
/rpt_family="ALU"
18667. .19235
/note="match to human fetal brain 5' EST D61494
(NID:9970409), bases 1-255, and to human 3' EST R07476
(NID:9759399)"
19550. .19670
/rpt_family="ALU"
21507. .37303
/note="similarity to various SS-RNA virus polyproteins;
pseudogene: region of matches and close matches to
multiple human ESTs, see R68740 (NID:9842257)"
37316. .37489
/note="Grail prediction, score = 80"
evidence=not experimental
complement(38938. .39224)
/rpt_family="ALU"
39225. .39707
/note="match to multiple human ESTs, see N30113
(NID:91148633)"
39800. .40085
/rpt_family="ALU"
complement(40247. .40538)
/rpt_family="ALU"
complement(40632. .40924)
/rpt_family="ALU"
complement(42283. .42891)
/rpt_family="ALU"
complement(45474. .45613)
/rpt_family="ALU"
complement(45614. .45737)
/note="match to human 3' EST H48998 (NID:9988738), bases
129-333"
complement(46107. .47026)
/note="match to multiple human ESTs, see N81064
(NID:91243765), H48897 (NID:9988737), and M78831
(NID:9273146)"
complement(47027. .47318)
/rpt_family="ALU"
complement(47365. .47782)
/note="match to multiple human ESTs, see W37495
(NID:91319089)"
47898. .48115
/note="match to human 5' EST H62306 (NID:91015138), bases
93-368"
complement(48116. .48405)
/rpt_family="ALU"
complement(48406. .48584)
/note="match to human 3' EST N29952 (NID:91148472), bases
290-455, and 5' EST R12730 (NID:9765806)"
complement(48787. .49405)
/rpt_family="ALU"
complement(49406. .49534)
/note="match to human 3' EST R65794 (NID:9838432), bases
309-440"

```

```

repeat_region      complement(49638..49672)
                    /rpt_family="ALU"
misc_feature        /note="match to human 3' EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148438), sequences are from opposite
                    ends of the same clone"
gene                /gene="WUSC:H.RG083M05.2"
                    /complement(49698..51806)
                    /gene="WUSC:H.RG083M05.2"
                    /complement(49698..51806)
                    /note="coded for by human cDNA #37388 (NID:g1319205),
                    R6591 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
CDS                 /codon_start=1
                    /protein_id="AAB46345.1"
                    /db_xref="GI:1669370"
                    /translation="MFYFPGGILFPCPGVYVQIGVSVDEQDKRYAALRGF
                    IQDYCEKSAALWMLPTLSPRIQDFDASTITGEEDLPKMEYLEFVCHAPSEYFK
                    SRSSPFVPVPTKGYIMTVGHTPALTIKESVANHL"
                    /complement(51576..51751)
                    /gene="WUSC:H.RG083M05.2"
                    /note="Grail prediction, score = 86"
exon                /evidence=not experimental
                    /complement(52052..52325)
                    /rpt_family="L1"
                    /note="match to human EST M79192 (NID:g273505) base 2-289"
repeat_region      55557..55843
misc_feature        /note="match to human EST M79192 (NID:g273505) base 2-289"

Query Match      99.4%; Score 498; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 7,2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 TCGGCAACCTCCCAACGACCTAGGTTTCCTTGAGATGGGGGAGCTAGAGACAG 63
Db 37504 TCGGCAACCTCCCAACGACCTAGGTTTCCTTGAGATGGGGGAGCTAGAGACAG 37563
Qy 64 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAGCTGAGAGTGAACAC 123
Db 37564 GACTAGCTGATTTCTCTAGGCTGACTAAGAAATCCCTAGCTGAGAGTGAACAC 37623
Qy 124 ATCCACCTTTAAACAGGGGCTTGCACTAGCTCAACCTGACCAATGAGAGCTCAC 183
Db 37624 ATCCACCTTTAAACAGGGGCTTGCACTAGCTCAACCTGACCAATGAGAGCTCAC 37683
Qy 184 TAAATGCTAATTAGGCAAGAAGAGGTTAAAGAAATAGCCAACTATTTGCTGAG 243
Db 37684 TAAATGCTAATTAGGCAAGAAGAGGTTAAAGAAATAGCCAACTATTTGCTGAG 37743
Qy 244 AGCAGACGAGAGGAGCAATGATCGGATATAAACCNAAGCTTCGAGCCGCAACGGCA 303
Db 37744 AGCAGACGAGAGGAGCAATGATCGGATATAAACCNAAGCTTCGAGCCGCAACGGCA 37803
Qy 304 ACCCCCTTTGGGTCCTCTCTTTGATGGAGCTCTTTTCATGCTATTTTCACTCTAT 363
Db 37804 ACCCCCTTTGGGTCCTCTCTTTGATGGAGCTCTTTTCATGCTATTTTCACTCTAT 37863
Qy 364 TAAATCTTGAACCTGACTCTTCTGGTCCATGTTTCTTAACGGCTGAGCTTTCG 423
Db 37864 TAAATCTTGAACCTGACTCTTCTGGTCCATGTTTCTTAACGGCTGAGCTTTCG 37923
Qy 424 TCGCATTCACCACTGCTGTTTGCAGCAACCGGACCCGCGCTGACTCCCATCCCTCT 483
Db 37924 TCGCATTCACCACTGCTGTTTGCAGCAACCGGACCCGCGCTGACTCCCATCCCTCT 37983
Qy 484 GGATCATGCAAGGCTGCC 501
Db 37984 GGATCATGCAAGGCTGCC 38001

```

```

RESULT 4
AC007566/c AC007566 149194 bp DNA linear PRI 01-MAR-2002
LOCUS      Homo sapiens BAC clone CTB-1065 from 7q21-7q22, complete sequence.
DEFINITION

```

```

ACCESSION          AC007566
VERSION            AC007566.2
KEYWORDS           GI:11181861
SOURCE             HTG.
ORGANISM           Homo sapiens.
                   Homo sapiens.
REFERENCE           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS            1 (bases 1 to 149194)
                   Sulston, J.E. and Waterston, R.
TITLE              Toward a complete human genome sequence
JOURNAL            Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE            99063792
PUBMED             9847074
REFERENCE           Du, Z.
                   2 (bases 1 to 149194)
TITLE              The sequence of Homo sapiens BAC clone CTB-1065
JOURNAL            Unpublished (2001)
REFERENCE           3 (bases 1 to 149194)
                   Waterston, R.H.
TITLE              Direct Submission
JOURNAL            Submitted (02-OCT-2000) Department of Genetics, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           4 (bases 1 to 149194)
                   Waterston, R.H.
TITLE              Direct Submission
JOURNAL            Submitted (06-FEB-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           5 (bases 1 to 149194)
                   Waterston, R.H.
TITLE              Direct Submission
JOURNAL            Submitted (16-NOV-2000) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           6 (bases 1 to 149194)
                   Waterston, R.H.
TITLE              Direct Submission
JOURNAL            Submitted (03-JAN-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           7 (bases 1 to 149194)
                   Waterston, R.H.
TITLE              Direct Submission
JOURNAL            Submitted (06-FEB-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE           8 (bases 1 to 149194)
                   Waterston, R.
TITLE              Direct Submission
JOURNAL            Submitted (01-MAR-2002) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
                   On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT            ----- Genome Center
                   Center: Washington University Genome Sequencing Center
                   Center code: WUGSC
                   Web site: http://genome.wustl.edu/gsc
                   Contact: saplens@watson.wustl.edu
                   ----- Summary Statistics
                   Center project name: H_RG010050

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/sec>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.reagen.com>).
VECTOR: pBelOAPC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```
1..149194
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /map="7q21-7q22"
  /clone_1lb="CTB-10G5"
  /clone_1lb="CTB-978SK-B"
  /tpc_family="L1"
  /note="match to EST BG752883 (NID:g14063536)"
  2248..2387
  /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
  2248..2387
  /note="match to EST AM579261 (NID:g7254310)"
  2248..2387
  /note="match to EST BG766882 (NID:g14077535)"
  2248..2387
  /note="match to EST B1160365 (NID:g14620366)"
  2248..2374
  /note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
  2248..2287
  /note="match to EST AV686676 (NID:g10288539)"
  2253..2387
  /note="match to EST AU123510 (NID:g10948226)"
  2344..2387
  /note="match to EST AV686676 (NID:g10288539)"
  2696..3066
  /tpc_family="MAlr"
  /tpc_family="Alu"
  3540..3628
  /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
  3540..3628
  /note="similar to Mus musculus EST BB253526 (NID:g8946272)"
  3542..3632
  /note="match to EST AV686676 (NID:g10288539)"
  3542..3628
  /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
  3542..3628
  /note="match to EST AU123510 (NID:g10948226)"
  3542..3628
  /note="match to EST BG766882 (NID:g14077535)"
```

```
misc_feature 3542..3628
             /note="match to EST B1160365 (NID:g14620366)"
misc_feature 3542..3609
             /note="match to EST AM579261 (NID:g7254310)"
misc_feature 3562..3628
             /note="match to EST BE272564 (NID:g9146913)"
misc_feature 3598..3628
             /note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714..3785
             /note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717..3785
             /note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719..3785
             /note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
             /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
misc_feature 3719..3785
             /note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
             /note="match to EST BE272564 (NID:g9146913)"
misc_feature 3719..3785
             /note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719..3785
             /note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719..3785
             /note="match to EST B1160365 (NID:g14620366)"
misc_feature 3719..3785
             /note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3860..3862
             /note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730..4827
             /tpc_family="L2"
             /tpc_family="L1"
             /note="match to EST AA425526 (NID:g2106267) zw48D03.r1"
             /note="match to EST AU123510 (NID:g10948226)"
             /note="match to EST BE272564 (NID:g9146913)"
             /note="match to EST BE299708 (NID:g9183456)"
             /note="match to EST BG260659 (NID:g12770475)"
             /note="match to EST BG766882 (NID:g14077535)"
             /note="match to EST B1160365 (NID:g14620366)"
             /note="similar to Mus musculus EST BE994936"
Query Match 99.4%; Score 498; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 8.2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	244	AGCACAGAGGAGGGAACAATGATCGGAAATATAAACCAAGTCTTGAGCCGGCAACGGCA	303
Db	84260	AGCACAGCAGGAGGAGACANTGATCGGGAATATAACCCAAGTCTTGAGCCGGCAACGGCA	84201
QY	304	ACCCCTTTGGGTCCCTCCCTCTTTGATGGAGGCTCTGTTTCATGCTAATTTCACTTAT	363
Db	84200	ACCCCTTTGGGTCCCTCCCTCTTTGATGGAGGCTCTGTTTCATGCTAATTTCACTTAT	84141
QY	364	TAAATCTTGCAACTGCACTCTTGAGTCACATTTATTAACGGCTGAGGTAGCTTTGGC	423
Db	84140	TAAATCTTGCAACTGCACTCTTGAGTCACATTTATTAACGGCTGAGGTAGCTTTGGC	84081
QY	424	TGCGCATCCACCACCTGCTGTGTTTGCCGCCACCGCAGACCCGCGCTGACTCCATCCCTCT	483
Db	84080	TGCGCATCCACCACCTGCTGTGTTTGCCGCCACCGCAGACCCGCGCTGACTCCATCCCTCT	84021
QY	484	GGATCATGCAAGGCTGCC	501
Db	84020	GGATCATGCAAGGCTGCC	84003

RESULT 5			
AX007998			
LOCUS	AX007998	711 bp	DNA
DEFINITION	Sequence 21 from Patent WO967395.		linear
ACCESSION	AX007998		
VERSION	AX007998.1	GI:9995695	
KEYWORDS			
SOURCE	human.		

FEATURES	REFERENCES
EUKARYOTA, METAZOA, CHORDATA, CRANIAL VERTEBRATA, EUCELEOSTOMI	Eukaryota, Metazoa; Chordata; Cranial Vertebrata; Euceleostomi
MAMMALIA, EUTHERIA, PRIMATES, CATARRHINI, HOMINIDAE, HOMO.	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
PERIN, J.P., RIEGER, F. and ALLIEL, P.M.	1 (bases 1 to 711)
Nucleic sequence and deduced protein sequence family with human	Perin, J.P., Rieger, F. and Alliel, P.M.
endogenous retroviral motifs, and their uses	Nucleic sequence and deduced protein sequence family with human
Patent: WO 9667395-A 21 29-DEC-1995	endogenous retroviral motifs, and their uses
INST NAT SANTRE RECH MED (FR), PEIXIN JEAN PIERRE (FR), RIEGER	Patent: WO 9667395-A 21 29-DEC-1995
FRANCOIS (FR), ALLIEL PATRICK M (FR)	INST NAT SANTRE RECH MED (FR), PEIXIN JEAN PIERRE (FR), RIEGER
Location/Qualifiers	FRANCOIS (FR), ALLIEL PATRICK M (FR)

FEATURES	Location/Qualifiers
source	1..711 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	171 a 209 c 160 g 171 t
ORIGIN	

Query Match 89.8%; Score 450; DI3 6; Length 711;

Best Local Similarity 100.0%; Pred. No. 1,16-134;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

52 ACTGAGAGACGAGACTAGCTGATTCCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGG 11

Db 1 ACTGAGACAGGACTAGCTGATTTCCTAGGCTGACTAAGAATCCCTAAGCCTAGCTGG 60

112 GAAGGTGACCACATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACACCTGACCAAT 17

Db 61 GAAGGTGACCACATCCACCCTTAAACACGGGGCTTGCACTTAGCTCACACCTGACCAAT 12

QY 172 CAGAGAGCTCACTAAATGCTAATTAGGCAAGACAGGAGGTAAAGAAATAGCCAAICAT 23

DB 121 CAGAGAGCTCACCTAAATGCTAATTAGGCATAAGACAGGAGGTAATGAAATAGCCAACTCAT 10

232 CTAATGCCCGAGGACACAGCAGGAGGCATAIGATGGAAATATATCCACCAGGTCCTCCAGG 23

PB 181 CIAIGC LGAGGCHLACAGGAGCACMAI GAI CGS . H . N . M . L CCGGGCCCAVCCAA

Z

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

Accession	Sequence	Position
Oy	CTGAGCTTTGGCTGGCCATCCACACAGCTGTTGGCCGACCGCAGACCGCGCTGAC	472
Oy	412	
Db	CTGAGCTTTGGCTGGCCATCCACACAGCTGTTGGCCGACCGCAGACCGCGCTGAC	422
Db	361	
Oy	TCCCATCCCTCTGAGTCATGAGGATGCC	501
Oy	472	
Db	TCCCATCCCTCTGAGTCATGAGGATGCC	450
Db	421	

RESULT 6	AP001538/c	LOCUS	DEFINITION
AP001538	174019 bp	DNA	linear
Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:BB53K11, L15.6-AP region, complete sequence.			

ACCESSION	AP001538
VERSION	AP001538.1
KEYWORDS	HTG.
SOURCE	Homo sapiens DNA, clon
ORGANISM	Homo sapiens

ORGANISM *nomio sapientis*
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE	1 (bases 1 to 174019)	Taylor, T. D., Hong-Seog, P.
AUTHORS	Hattori, M., Ishii, K., Toyoda, A.,	Watanabe, H. and Sakaki, Y.
	Fujiyama, A., Yada, T.,	Totoki, Y.

TITLE Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2
JOURNAL Published Only in Database (2000)

REFERENCE	2 (cases 1 to 1/4019)	AUTHORS
Hattori, M., Ishii, K., Fujiyama, A., Yada, T., Tsuchi, Y.,	Toyoda, A., Taylor, T. D., Hong-Seog, P. Watanabe, H. and Sakaki, Y	

TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa
Japan (E-mail: hattori@sc.riken.go.jp)

FEATURES
Location/Qualifiers
URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-42-778-9923
Fax: 81-42-778-9924

```
source
1. .174019
/organism="Homo sapiens"
/ab_xref="taxon:9606"
```

```
/wd xref= cavi01:5000  
/chromosome="21"  
/map="21q21.1-q21.2"
```

BASE COUNT	53303	a	31416	c	32248	g	57052	t
ORIGIN	/clone="B853K11"							

Query Match	Score	DB 9	Length
86.6%	434	9	174019
86.8%	4	100	

Best Local Similarity 92.0%; P-Seq. NO. 4,354,125
Matches 456; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

4 TCGGCCAAGCTCCCAACAGACACTTAGTTTCTGTGAGATGGGGGACTGAGAGACAG 63

54 GACTAGCTGATTCTTAGGCTGACTAAGATCCCTAAGCTAGCTGGGAAGTGAACAC 12

Db 36408 GACTAGCTGGATTCTCCTAGGCTGACTAAGAATCCCTAAGCTAGCTGGGAAGTGACCGC 36

QY 124 ATCCACCTTAAACAGGGGCTTGAACCTTAGCTCACACCTGACCAATCAGAGACTCAC 18

Db 36348 ATCCACTTTAAACAATGGGGCTCCGCACTTAGCTACACCCCAACAGTACAGAGAGCTCAC 36

104 TAAATATGCTAATAGGCAAAACAGAGGTAAAGAAATAGCCATCATCTATGCGCTGAG 36288
 36288 Db

244 AGCAGCAGGAGGCAATGATCGGATATTAACCCAGTCTTCGAGCCGCGCAACGGCA 30

Db 36228 AGCAGCAGGAGGGACAAAGATCAGATGTTAAACCGAGCAATTCAAGCTGGCAACGGCA 36

304 ACCCCCCCGGGCCCCCCCCCCCCCCCCCCAGGAGCAGCCGACCTTGTATTTTCAGTAACTGCCTTCA
 |||||
 |||||

[illegible]

```

/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R28M9, 3' partial"
/clone_id="RRC1-11 BAC library"
/note="Accession No. AP000431"
311751. .3140000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="c103E2331, 3' partial"
/clone_id="ICRFc103 Coe mid library"
/note="Accession No. AL132933"
15. .444
/note="L1MA4A"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
555. .3531
/note="L1MA4A"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
3770. .3876
/note="L2"
/rpc_family="LINE/L2"
/rpc_type=DISPERSED
5645. .6178
/note="L1MA2"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
6181. .6534
/note="THE1B"
/rpc_family="LTR/MaLR"
/rpc_type=DISPERSED
8021. .8370
/note="THE1B"
/rpc_family="LTR/MaLR"
/rpc_type=DISPERSED
8373. .10032
/note="L1MA2"
/rpc_family="LINE/L1"
/rpc_type=DISPERSED
10213. .10520
/note="AluDb"
/rpc_family="SINE/Alu"
/rpc_type=DISPERSED
15810. .15851
/note="(CA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
complement(16628. .16697)
/note="MER53"
/rpc_family="DNA"
/rpc_type=DISPERSED
16638. .16715
/note="MER53"
/rpc_family="DNA"
/rpc_type=DISPERSED
18847. .19585
/note="L2"
/rpc_family="LINE/L2"
/rpc_type=DISPERSED
19735. .19917
/note="(TA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
20023. .20166
/note="(TA)n"
/rpc_family="Simple_repeat"
/rpc_type=TANDEM
20285. .20448

```

Query Match 86.6%; Score 434; DB 9; Length 340000;

Best Local Similarity 92.0%; Pred. No. 5,4e-129;
Matches 458; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

Qy 4 TCGGCCAACCTCCCAACAGACCTTAGGTTTCTGTTGAGATGGGAGCTGAGAGACAG 63
Db 27679 TTGGCCAACTCCCAACAGACCTTAGGTTTCTGTTGAGAGGCGGACTGAGAGACAG 27620
Qy 64 GACTAGCTGATTTCTGAGCTGATCAAGAAATCCCTAAGCTAGCTGGAGAGTACAC 123
Db 27619 GACTAGCTGATTTCTGAGCTGATCAAGAAATCCCTAAGCTAGCTGGAGAGTACAC 27560
Qy 124 ATCCACTTTAAACAGGCGCTTGAATAGCTCAGACCTGACCAATCAAGAGCTCAC 183
Db 27559 ATCCACTTTAAACAGGCGCTTGAATAGCTCAGACCTGACCAATCAAGAGCTCAC 27500
Qy 184 TAAATGCTAATAGGCAAGACAGAGATTAAGAAATAGCAATCATCTATTGCTCAG 243
Db 27439 TAAATGCTAATAGGCAAGAAACAGAGATTAAGAAATAGCAATCATCTATTGCTCAG 27440
Qy 244 AGCAGCAGAGAGGAGCAATATGCGGATTAACCCAGTCTTGAGCCGCGCAAGCA 303
Db 27439 AGCAGCAGAGAGGAGCAATATGCGGATTAACCCAGTCTTGAGCCGCGCAAGCA 27380
Qy 304 ACCCCCTTGGGTCCCTCCCTTTGTATGAGGAGCTGTTTTCATGCTATTCTACTAT 363
Db 27379 ACCCCCTTGGGTCCCTCCCTTTGTATGAGGAGCTGTTTTCATGCTATTCTACTAT 27320
Qy 364 TAAATCTGCAACTGCACTCTTCTGTCATGTTCTTCTTCAAGCTGAGCTGCTTGC 423
Db 27319 TAAATCTGCAACTGCACTCTTCTGTCATGTTCTTCTTCAAGCTGAGCTGCTTGC 27260
Qy 424 TCGCCATCCACCACTGCTGTTTCCGCGCACCGACCCGCGCTGACTCCCATCTCT 483
Db 27259 TCGCTTCCACCACTGCTGTTTCCGCGCGAGACCTGCGACACTGCCATCTCCCTCC 27200
Qy 484 GGATCATGAGGCTGCTCC 501
Db 27199 GGATCGGAGGCTGCTCC 27182

```

RESULT 8
AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
AC093531
AC093531.2 GI:16945981
VERSION
KEYWORDS
HTG.
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 163803)
REFERENCE
TITLE
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL
Unpublished
2 (bases 1 to 163803)
REFERENCE
TITLE
DOE Joint Genome Institute.
JOURNAL
Direct Submission
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 163803)
REFERENCE
TITLE
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL
Direct Submission
Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Nov 16, 2001 this sequence version replaced gi:15383820.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.
Location/Qualifiers
FEATURES

source 1. .163803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"

BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN

Query Match 86.1%; Score 431.2; DB 9; Length 163803;
Best Local Similarity 91.4%; Pred. No. 4e-128;
Matches 457; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2 TCTGGCCAACTCCCAAGACGACCTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 61
Db 117108 TGTCAAGCAACTCCCAAGACGACCTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 117049

QY 62 AGACTAGCTGATTTCTAGCTGACCTAAGATCCCTAAGCTAGCTGGGAAGTACC 121
Db 117048 AGACTAGCTGATTTCTAGCTGACCTAAGATCCCTAAGCTAGCTGGGAAGTACC 116989

QY 122 ACATCACTTTAAACAGGGGCTTGCACCTTACCTGACCTGACCTGACCTG 181
Db 116988 GCATCCATCTTTAAACAGGGGCTTGCACCTTACCTGACCTGACCTGACCTG 116929

QY 182 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCAATCATCTATTGCTG 241
Db 116928 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCAATCATCTATTGCTG 116869

QY 242 AGAGCAGAGAGAGAGAGCAATGATGGGATTAACCCCAAGTTTGGAGCGGCAACGG 301
Db 116868 AGAGCAGAGAGAGAGAGCAATGATGGGATTAACCCCAAGTTTGGAGCGGCAACAG 116809

QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCACTATTCACTCT 361
Db 116808 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCACTATTCACTCT 116749

QY 362 ATTAATCTTGCACTGACCTCTCTGCTCATGTTTCTTAAGCTTGAAGTGAAGCTTC 421
Db 116748 ATTAATCTTGCACTGACCTCTCTGCTCATGTTTCTTAAGCTTGAAGTGAAGCTTC 116689

QY 422 GCTGGCATCAGACGCTGCTGTTGGCCGACCGGACCGCGGCTGACCTCCATCCCT 481
Db 116688 GCTGGCATCAGACGCTGCTGTTGGCCGACCGGACCGCGGCTGACCTCCATCCCT 116629

QY 482 CTGGATCATGAGGGTGTC 501
Db 116628 TTGGATCCAGCAGAGTGTC 116609

RESULT 9
AF127226 839 bp DNA linear PRI 17-JUN-1999
LOCUS AF127226
DEFINITION Homo sapiens Human endogenous retrovirus W, 3' long terminal
repeat, partial sequence.
ACCESSION AF127226
VERSION AF127226.1 GI:5081468
KEYWORDS
SOURCE Homo sapiens
ORGANISM
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 839)
Ounanian-Paraz, A., Komurian-Pradel, F., Ott, C., Rajorharrison, A. and
Perron, H.
TITLE MSRV-related HERV-W elements in human DNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 839)
AUTHORS Ounanian-Paraz, A., Komurian-Pradel, F., Ott, C., Rajorharrison, A. and
Perron, H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1999) UMR103 CNRS-Biomerieux, Biomerieux, 46,
Allee D' Italie, Lyon 69007, France
FEATURES
Location/Qualifiers

source 1. .839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="3' long terminal repeat; U3, R and U4 regions"

BASE COUNT 205 a 233 c 192 g 206 t 3 others
ORIGIN

Query Match 85.0%; Score 426; DB 9; Length 839;
Best Local Similarity 90.6%; Pred. No. 9.9e-127;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCAACTCCCAAGACGACCTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 61
Db 4 TGTCAAGCAACTCCCAAGACGACCTTAAGTTTCTGTTGAGTGGGGAGCTAGAGAC 63

QY 62 AGACTAGCTGATTTCTAGCTGACCTAAGATCCCTAAGCTAGCTGGGAAGTACC 121
Db 64 AGACTAGCTGATTTCTAGCTGACCTAAGATCCCTAAGCTAGCTGGGAAGTACC 123

QY 122 ACATCACTTTAAACAGGGGCTTGCACCTTACCTGACCTGACCTGACCTG 181
Db 124 GCATCCATCTTTAAACAGGGGCTTGCACCTTACCTGACCTGACCTGACCTG 183

QY 182 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCAATCATCTATTGCTG 241
Db 184 ACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCAATCATCTATTGCTG 243

QY 242 AGAGCAGAGAGAGAGAGCAATGATGGGATTAACCCCAAGTTTGGAGCGGCAACGG 301
Db 244 AGAGCAGAGAGAGAGAGCAATGATGGGATTAACCCCAAGTTTGGAGCGGCAACAG 303

QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCACTATTCACTCT 361
Db 304 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCACTATTCACTCT 363

QY 362 ATTAATCTTGCACTGACCTCTCTGCTCATGTTTCTTAAGCTTGAAGTGAAGCTTC 421
Db 364 ATTAATCTTGCACTGACCTCTCTGCTCATGTTTCTTAAGCTTGAAGTGAAGCTTC 423

QY 422 GCTGGCATCAGACGCTGCTGTTGGCCGACCGGACCGCGGCTGACCTCCATCCCT 481
Db 424 GCTGGCATCAGACGCTGCTGTTGGCCGACCGGACCGCGGCTGACCTCCATCCCT 483

QY 482 CTGGATCATGAGGGTGTC 501
Db 484 TTGGATCCAGCAGAGTGTC 503

RESULT 10
AX001030 1329 bp DNA linear PAT 10-MAR-2000
LOCUS AX001030
DEFINITION Sequence 12 from Patent WO9902666.
ACCESSION AX001030
VERSION AX001030.1 GI:7241266
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 1329)
AUTHORS Ott, C. and Bedin, F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL PATENT: WO 9902666-A 12 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES
Location/Qualifiers
1. .1329
/organism="unidentified"

BASE COUNT 346 a /db_xref="taxon:32644"
 ORIGIN 352 c 288 g 340 t 3 others

Query Match 85.0%; Score 426; DB 6; Length 1329;
 Best Local Similarity 90.6%; Pred. No. 1=126;
 Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGGCAACCTCCCAAGACGACCTTAGTGTTCCTGTTGAGATGGGGGAGCTGAGAGAC 61
 DB 497 TGTCAACCAACCTCCCAAGACGACCTTAGTGTTCCTGTTGAGAGATGGGGGAGCTGAGAGAC 556

QY 62 AGAAGTCTGATTTCTTAGGCTGACCTAAGATCCTAAGCTTAGCTGGAGAGCTGAC 121
 DB 557 AGAAGTCTGATTTCTTAGGCTGACCTAAGATCCTAAGCTTAGCTGGAGAGCTGAC 616

QY 122 ACATCCACTTTAAACAGGGGGCTTGCAACTTAGCTGACCTGACCAATGAGAGCTG 181
 DB 617 GCATCCACTTTAAACAGGGGGCTTGCAACTTAGCTGACCTGACCAATGAGAGCTG 676

QY 182 ACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCCATGCTATTTGCTG 241
 DB 677 ACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCCATGCTATTTGCTG 736

QY 242 AGAGCAGCAGAGAGAGGCAATGATCGGATATTAATCCAACTTTCAGCCGCAACGG 301
 DB 737 AGAGCAGCAGAGAGAGGCAATGATCGGATATTAATCCAACTTTCAGCCGCAACGG 796

QY 302 CAACCCCTTTGGGCTCCCTCCTTTGATAGGAGCCCTGTTTATGCTATTTGCTG 361
 DB 797 CAACCCCTTTGGGCTCCCTCCTTTGATAGGAGCCCTGTTTATGCTATTTGCTG 856

QY 362 ATTAATCTTGCACTGCACTCTTCTGCTCATGTTTCTTAAGGCTTGAAGCTTTC 421
 DB 857 ATTAATCTTGCACTGCACTCTTCTGCTCATGTTTCTTAAGGCTTGAAGCTTTC 916

QY 422 GCTGCGCATCCACCATGCTGTTTCCCGCACGACACCGCCGCTGATCCCATCCCT 481
 DB 917 GCTGCGCATCCACCATGCTGTTTCCCGCACGACACCGCCGCTGATCCCATCCCT 976

QY 482 CTGATCATGCAAGGTGTC 501
 DB 977 TTGATCCAGCAGAGGTGTC 996

RESULT 11
 AX000970 783 bp 1MA linear PAT 10-MAR-2000
 LOCUS AX000970
 DEFINITION Sequence 15 from Patent WO9902696.
 ACCESSION AX000970
 VERSION AX000970.1 GI:7241212
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 783)
 AUTHORS Beseme, F. and Blond, J.
 TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
 DISEASES OR WITH PREGNANCY DISORDERS
 JOURNAL Patent: WO 9902696-A 15 21-JAN-1999;
 BIO MEDICX (FR); BESEME FREDERIC (FR)
 FEATURES
 source 1..783
 location/Qualifiers
 /organism="unidentified"
 /db_xref="taxon:32644"

BASE COUNT 173 a 213 c 166 g 180 t 51 others
 ORIGIN

Query Match 85.0%; Score 425.6; DB 6; Length 783;
 Best Local Similarity 94.0%; Pred. No. 1.3e-126;
 Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;

QY 54 TGAGAGACAGACTAGTGTATTTCTTAGCTGACTGACTAGCAATCCCTTAAGCTTAGCTGGGA 113

DB 1 TGAGAGACAGACTAGTGTATTTCTTAGCTGACTGACTAGCAATCCCTTAAGCTTAGCTGGGA 60

QY 114 AGGTGACCAATCCACTTTAAACACGGGGCTTGCAACTTAGCTGACACTGACCAATCA 173
 DB 61 AGGTGACCAATCCACTTTAAACACGGGGCTTGCAACTTAGCTGACACTGACCAATCA 120

QY 174 GAGAGCTCACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATTAACCAATCACT 233
 DB 121 GAGAGCTCACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATTAACCAATCACT 180

QY 234 ATTGCTGAGAGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGCTTTCAGACC 293
 DB 181 ATTGCTGAGAGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGCTTTCAGACC 240

QY 294 GGCACGCA-ACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTTTCATGCTA 352
 DB 241 GGCACGCAAGCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTTTCATGCTA 300

QY 353 TTTGACTATTAATCTTGAATGCACTCTTGTGCTCACTGTTTCTTAAGGCTTTCAGC 412
 DB 301 TTTGACTATTAATCTTGAATGCACTCTTGTGCTCACTGTTTCTTAAGGCTTTCAGC 360

QY 413 TGAGCTTTCGCTGCGCATCCACCACTGCTGTTTGCAGCAGCCGCAACCGCCGCTGACT 472
 DB 361 TGAGCTTTCGCTGCGCATCCACCACTGCTGTTTGCAGCAGCCGCAACCGCCGCTGACT 420

QY 473 CCCATCCCTTGATCATGCAAGGTGTC 501
 DB 421 CCCATCCCTTGATCATGCAAGGTGTC 449

RESULT 12
 HUAE000660 251124 bp DNA linear PRI 15-JUN-2001
 LOCUS HUAE000660
 DEFINITION Homo sapiens T-cell receptor alpha delta locus from bases 501613 to 752736 (section 3 of 5) of the Complete Nucleotide Sequence.
 ACCESSION AB000660 AB000521 U85197
 VERSION AB000660.1 GI:2358042
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 251124)
 AUTHORS Koop, B.F., Rowen, L., Wang, K., Xuo, C.L., Seto, D., Lemstra, J.A., Howard, S., Shan, W., Deshpande, P. and Hood, L.
 TITLE The human T-cell receptor TCRAC/TCRDC (C alpha/C delta) region: organization, sequence, and evolution of 97.6 kb of DNA
 JOURNAL Genomics 19 (3), 478-493 (1994)
 MEDLINE 94245236
 PUBMED 8188290
 REFERENCE 2 (bases 1 to 251124)
 AUTHORS Boyesen, C., Simon, M.I. and Hood, L.
 TITLE Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones
 JOURNAL Genome Res. 7 (4), 330-338 (1997)
 MEDLINE 97264339
 PUBMED 9110172
 REFERENCE 3 (bases 1 to 251124)
 AUTHORS Boyesen, C., Inyoun, L., Smith, T.M., Smit, A., Wang, K., Rowen, L. and Hood, L.
 TITLE T-cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
 JOURNAL Unpublished
 REFERENCE 4 (bases 1 to 251124)
 AUTHORS Boyesen, C., Lee, I., Smith, T.M., Smit, A., Kai, W., Lee, R. and Leroy, H.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-1997) Department of Molecular Biotechnology, University of Washington, Box 357330, Seattle, Washington 98195, USA
 REMARK Complete nucleotide sequence of the human T-cell receptor alpha

COMMENT

delta locus
On Jun 12, 2001 this sequence version replaced gi:2121231.
Sequencing method: high redundancy shotgun. Interspersed Repeats
were identified with RepeatMasker (available from
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple
sequence repeats were identified with sputnik (available from
<http://serac.mbt.washington.edu/~chrisa/software/sputnik.html>).
Location/Qualifiers

FEATURES

```
source
1..251124
/organism="Homo sapiens"
/db_xref="taxon:9606"
157356..235794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BAC 480"
/clone_lib="Mel Simon's BAC library"
/note="This span of sequence was submitted independently
as Genbank Accession Number U85197."
235792..>251124
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BAC 378"
/clone_lib="Mel Simon's BAC library"
/note="This span of sequence was submitted independently
as Genbank Accession Number U85196."
/rpt_family="MIR"
603..983
/rpt_family="MLT1B"
1738..2478
/gene="TCRAV17S1"
1738..1757
/standard_name="TCRAV17S1"
/note="Old name TCRAV17S1"
join(<1957..2008,2155..>2439)
/gene="TCRAV17S1"
/product="TCRAV17S1"
/note="Old name TCRAV3S1"
join(1957..2008,2155..>2439)
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
/codon_start=1
/product="TCRAV17S1"
/protein_id="AAB69018.1"
/db_xref="GI:2358043"
/translation="MTLLGVSLVILMLQLARVNSQGEEDPOLSTIOEGENATMNC
YKTSINNLQWYRQNSGRGLVHLILRSNEREKSGRLVTLDTSKSSILLTVASRAA
DTASVFCATD"
join(1957..2008,2155..>2439)
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
2440..2446
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="heptamer"
2447..2469
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="spacer"
2470..2478
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="nonamer"
complement(4306..4461)
/rpt_family="MIR2"
4975..5275
/rpt_family="AluY"
complement(5366..5574)
/rpt_family="MERS8A"
7244..7983
gene
```

```
/gene="TCRAV18S1"
/pseudo
7244..7263
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/pseudo
join(<7482..7527,7656..>7944)
/gene="TCRAV18S1"
/product="TCRAV18S1"
/pseudo
join(7482..7527,7656..>7944)
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/pseudo
7945..7951
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="heptamer"
7952..7974
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="spacer"
7975..7983
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="nonamer"
complement(8236..8389)
/rpt_family="MER69A"
complement(8556..8740)
/rpt_family="L1MB7"
complement(9281..10170)
/rpt_family="L1PA2"
complement(10023..11230)
/rpt_family="L1"
11234..11405
/rpt_family="L1"
11818..12491
/gene="TCRAV19S1"
11818..11837
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
join(<11900..11948,12152..>12452)
/gene="TCRAV19S1"
/product="TCRAV19S1"
/note="Old name TCRAV12S1"
join(11900..11948,12152..>12452)
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
/codon_start=1
/product="TCRAV19S1"
/protein_id="AAB69019.1"
/db_xref="GI:2358044"
/translation="MTASLRLAVIASICVSSMAQKVTQACTEISVEREKDYLDCV
VETRDITLYLFWYKPPSGELVFLIRNSPDEQNETSGRYSNMFGKSTSFNFTLAS
QVDSAVVFCALSEA"
join(11900..11948,12152..>12452)
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
12453..12459
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="heptamer"
12460..12482
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="spacer"
12483..12491
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="nonamer"
```

Query Match	83.0%	Score 415.6	DB 9	Length 251124
Best Local Similarity	90.5%	Pred. No. 4.9e-123		
Matches 458	Conservative 0	Mismatches 39	Indels 9	Gaps 1
repeat_region	/rpt_family="MIR2"	13763..13833		
repeat_region	/rpt_family="MIR"	13879..13946		
repeat_region	complement(19382..19671)			
repeat_region	/rpt_family="AluBg"	21487..21622		
repeat_region	/rpt_family="MIR"	23034..23706		
repeat_region	/rpt_family="JME3A"	23955..24354		
repeat_region	/rpt_family="WSTA"	24354..24523		
repeat_region	complement(24364..24523)			
repeat_region	/rpt_family="MIR2"	25045..25045		
Db 247322	TCAGCAACCTCCCAACAGCACTTAGTGTTCCTTGAGATGGGGGACGTGAGACAG	63		
Db 247322	GACTAGCTGGATTTCTTCTAGCTGACTGACCTTAACATCCCTTAAGCTAGCTGGAGAGTACCGC	123		
Db 247322	GACTAGCTGGATTTCTTCTAGCTGACTGACCTTAACATCCCTTAAGCTAGCTGGAGAGTACCGC	247381		
Db 247382	ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAACCTGACCAATC-----AG	174		
Db 247382	TTCCACCTTTAAACACGGGGCTTGCACTTAGCTCAACCTGACCAATC-----AG	247441		
Db 247442	AGAGCTCACTAAATGCTTAATAGGCAAAAGACAGAGAGTAAAGAAATAGCCATATCTTA	234		
Db 247442	AGAGCTCACTAAATGCTTAATAGGCTTAATAGGCAAAAGACAGAGAGTAAAGAAATAGCCATATCTTA	247501		
Db 247502	TTGCGTGAAGACAGACAGAGAGCAATGATCGGGATTAACCCAGATCTTCGAGCGC	294		
Db 247502	TCACCTGAAGACAGACAGAGAGCAATGATCGGGATTAACCCAGATCTTCGAGCGC	247561		
Db 247562	GCAAGGGCAACCCCTTTGGGTCCCTCCCTTGTATGGAGCTGTGTTTCATCTATT	354		
Db 247562	GCAAGGGCAACCCCTTTGGGTCCCTCCCTTGTATGGAGCTGTGTTTCATCTATT	247621		
Db 247622	TCACCTTAATTAATCTTGCACTGCACTTTGTGTCATGTTTCTTACAGGCTTGAGCTG	414		
Db 247622	TCACCTTAATTAATCTTGCACTGCACTTTGTGTCATGTTTCTTACAGGCTTGAGCTG	247681		
Db 247682	AGCTTTCGTCACCAATCCACCTGCTGTTTCTGCGCACCAGAGACCCGCGCTGACTCC	474		
Db 247682	AGCTTTCGTCACCAATCCACCTGCTGTTTCTGCGCACCAGAGACCCGCGCTGACTCC	247741		
Db 247742	CATCCCTCAAGATCCGGGAGGGGTGTC	247767		
RESULT 13	AC092510	187321 bp	DNA	linear HTG 03-AUG-2002
LOCUS	AC092510/c			
DEFINITION	Papio cynocephalus anubis clone RP41-107F15, WORKING DRAFT			
ACCESSION	AC092510			
VERSION	AC092510.2	GI:22094320		
KEYWORDS	HTG; HTGS PHASE2; HTGS_DRAFT.			
SOURCE	olive baboon.			
ORGANISM	Papio cynocephalus anubis			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio.			
AUTHORS	1 (bases 1 to 187321) Akhter, N., Antonellis, A., Ayele, K., Eckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C.,			

BROOKS, S., DIETRICH, N. L., GRANITE, S., GUAN, X., GUPTA, J.,
 HANSEN, N., HO, S.-L., IDOL, J. R., KARLINS, E., LARIC, P.,
 LEACH, S.-O., LEASAP, R., MADURO, O. L., MADURO, V. B.,
 MASIELLO, C., MASIELLO, C., MASKERI, B., MASTRIAN, S. D.,
 MCGLOTHLIN, J. C., McDOWELL, J., PAGUIRIGAN, C., PARSON, R.,
 PORTNOY, M. E., PRAESAD, A., SCHNEIDER, M. G., SCANTIRPO, S., THOMAS, J. W.,
 THOMAS, P. J., TOUCHMAN, J. W., TAUSSON, C., VOGT, J. L., WALKER, M. A.,
 WETHEBY, K. D., WIGGINS, L., YOUNG, A., ZHANG, L.-H. and GREEN, E. D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2. (bases 1 to 187321)
 Green, E. D.
 Direct Submission
 Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717
 Groveomont Circle, Gaitherburg, MD 20877, USA
 3. (bases 1 to 187321)
 Green, E. D.
 Direct Submission
 Submitted (03-AUG-2002) NIH Intramural Sequencing Center, 8717
 Groveomont Circle, Gaitherburg, MD 20877, USA
 On Aug 3, 2002 this sequence version replaced gi:14717323.

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in 920 bases and has been reviewed to rule out gross misassemblies; the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

Center NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_zoo@ngri.nih.gov
----- Project Information
Center project name: amf
Center clone name: 107F15

```

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big_Dye; 100% of reads
Assembly program: Pirap; version 0.990319
Consensus quality: 185897 bases at least Q40
Consensus quality: 186555 bases at least Q30
Consensus quality: 186768 bases at least Q20
Insert size: 165000; agarose-fp
Insert size: 149000; pulse-field-gel
Insert size: 186821; sum-of-contigs
Quality coverage: 12.3x in Q20 bases; agarose-fp
Quality coverage: 13.64x in Q20 bases; pulse-field-gel
Quality coverage: 10.87x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the subcloner.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 25945: contig of 25945 bp in length
25946 26045: gap of unknown length
26046 59962: contig of 33917 bp in length
59963 60062: gap of unknown length
60063 77104: contig of 17042 bp in length
77105 77204: gap of unknown length
77205 160478: contig of 83274 bp in length
160479 160578: gap of unknown length

```


The clone sequenced to the left is RP11-512N1, 2000 bp overlap; the clone sequenced to the right is RP11-656O23. Actual start of this clone is at base position 64513 of RP11-512N1, actual end is at base position 83412 of RP11-41904.

Data from AC068520 was used to finish the clone, AC092843.

```
FEATURES
SOURCE
    location/Qualifiers
    1..83412
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2"
    /clone="RP11-41904"
    /clone_id="RPC1-11"
    251..428
    /rpt_family="MIR"
    repeat_region
    /rpt_family="MIR"
    434..653
    /rpt_family="MER1_type"
    repeat_region
    885..999
    /rpt_family="Alu"
    repeat_region
    1382..1416
    /rpt_family="CATATA)n"
    repeat_region
    3189..3347
    /rpt_family="MER1_type"
    repeat_region
    3745..4097
    /rpt_family="MALR"
    repeat_region
    5089..5311
    /rpt_family="MIR"
    repeat_region
    6018..6430
    /rpt_family="ERV1"
    6441..6634
    /note="match to EST BE182435 (NID:g8661671)"
    misc_feature
    7447..7507
    /rpt_family="ERV1"
    repeat_region
    7486..7584
    /rpt_family="TAGA)n"
    repeat_region
    7566..7962
    /rpt_family="ERV1"
    repeat_region
    8392..8422
    /rpt_family="(TTTG)n"
    repeat_region
    8624..8722
    /rpt_family="MER103"
    repeat_region
    8837..8984
    /rpt_family="CR1"
    repeat_region
    9090..9839
    /rpt_family="ERV1"
    repeat_region
    10545..10621
    /rpt_family="GA-rich"
    repeat_region
    10557..10583
    /note="similar to Homo sapiens EST BE264057 (NID:g9137607)"
    misc_feature
    15765..15803
    /rpt_family="ERV1"
    repeat_region
    15794..16085
    /rpt_family="ERV1"
    repeat_region
    16086..16863
    /rpt_family="ERV1"
    repeat_region
    16935..17592
    /rpt_family="CR1"
    repeat_region
    17960..18105
    /rpt_family="L2"
    repeat_region
    18346..18422
    /rpt_family="MIR"
    repeat_region
    18526..18857
    /rpt_family="MALR"
    repeat_region
    18858..19224
    /rpt_family="MALR"
    repeat_region
    18915..18935
    /rpt_family="(A)n"
    repeat_region
    19225..19248
    /rpt_family="MALR"
    repeat_region
    19311..19373
    /rpt_family="L1"
    repeat_region
```

```
repeat_region 19387..19561
/rpt_family="L1"
repeat_region 20213..20233
/rpt_family="AT-rich"
repeat_region 20763..20932
/rpt_family="MIR"
repeat_region 20980..21353
/rpt_family="MALR"
repeat_region 21489..21589
/rpt_family="L2"
repeat_region 21668..21783
/rpt_family="MIR"
repeat_region 22245..22318
/rpt_family="MIR"
repeat_region 23790..23942
/rpt_family="L2"
repeat_region 24196..24218
/rpt_family="AT-rich"
repeat_region 24445..24563
/rpt_family="L1"
repeat_region 24564..24748
/rpt_family="MER1_type"
repeat_region 24749..25103
/rpt_family="L1"
repeat_region 25113..25236
/rpt_family="L1"
repeat_region 28015..28291
/rpt_family="Alu"
repeat_region 31569..33238
/rpt_family="L1"
repeat_region 33626..33996
/rpt_family="MALR"
repeat_region 34390..34670
/rpt_family="L1"
repeat_region 38082..38216
/rpt_family="T-rich"
repeat_region 40015..40036
/rpt_family="(TTTA)n"
repeat_region 40025..40800
/rpt_family="L1"
repeat_region 40803..41831
/rpt_family="L1"
misc_feature 41230..41235
/note="similar to Homo sapiens EST AA852771 (NID:g2941364)"
repeat_region 41829..42643
/rpt_family="L1"
repeat_region 42648..42832
/rpt_family="L1"
repeat_region 42833..43136
/rpt_family="Alu"
```

Query Match 81.6%; Score 409; DB 9; Length 83412;
Best Local Similarity 91.4%; Pred. No 5; Be-121; Matches 456; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

```
QY 4 TCGGCAACCTCCCAACAGCATTAGCTTTCTGTTGAGATGGGGACTGAGACAG 63
Db 16036 TCAGCAACCTCCCAACAGCATTAGCTTTCTGTTGAGATGGGGACTGAGACAG 16095
QY 64 GACTAGCTGATTTCTTGCTGCTGACTAAGATCCCTAAGCTAGTGGGAAGCTACAC 123
Db 16096 GACTAGCTGATTTCTTGCTGCTGACTAAGATCCCTAAGCTAGTGGGAAGCTACAC 16155
QY 124 ATCCACCTTTAAACACGGGGCTGCACTTGTGCACACTGACCAATCAGAGACTCAC 183
Db 16156 ATCCACCTTTAAATCGGGCTTGCACCTTGTGCACACTGACCAATCAGAGACTCGT 16215
QY 184 TAAATGCTAATTAGGCAAGACAGAGATTAAGAAATAGCAATCATCTATTGCTTGA 243
Db 16216 TAAATGCTAATTAGGCAAGAGAGATTAAGAAATAGCAATCATCTATTGCTTGA 16275
QY 244 AGCAGAGAGAGGAGCAATGATCGGGATATTAACCAATGCTTTGAGCCGGCAACGGCA 303
```


	Db	16276	AGCACAAGGAGGGACAAAGATTGGAGATATATCCAGGCATTGGAGTGGCAACAGA	16335
OY	304		ACCCTTTGGGTGCCCTCCCTTTTGATGGAGACTCTGTTTTCACTATTTACTTAT	363
Db	16336		ACCCCTTTGGGTGCCCTCCCTTTTGATGGAG--CTGTTTTCACTATTTCACTAT	16393
OY	364		TAAATCTTGCAACTGCACTCTTCTGTCATGTTTTCTTAAGGCTTAGCTGACTTGGC	423
Db	16394		TAAATCTTGCAACTGCACTCTTCTGTCATGTTTTCTTAAGGCTTAGCTGAACTTTCAC	16453
OY	424		TGCGCATCACACACGCMG--TTTGGCGCACCGGACGCCGGCGTGCATCCATCCCTC	482
Db	16454		TGCGCATCACACACGCTGTTTTGCGCGCGTGCAGACCCACTGCTGACTTCATTCTTC	16513
OY	483		TGGATCANGCAGGGGTGCC	501
Db	16514		TGGATCANGCAGGGGTGCC	16532
RESULT 15				
AL139038				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

AL139038 140756 bp DNA linear PRI 15-JUL-2001
 Human DNA sequence from clone RP11-456B18 on chromosome 13,
 complete sequence.
 AL139038
 AL139038.18 GI:14800148
 HTG.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 140756)
 Dunn,M.
 Direct Submission
 Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Jul 17, 2001 this sequence version replaced gi:13751261.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TREMBL; Wp:, WormPEP; information on the WormPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 13, constructed by the Sanger Centre Chromosome 13
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
 RP11-456B18 is from the library RPCI-11.2 constructed by the group
 of Pister de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
 RP11-456B18 it may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true right end of clone RP11-456B18 is at 140756 in this
 sequence. The true left end of clone RP11-7816 is at 66441 in this
 sequence. The true right end of clone RP11-108H9 is at 100 in this
 sequence.

Location/Qualifiers

source	1..140756	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
	/chromosome="13"	
	/clone="RP11-456b18"	
	/clone_1ib="RPC1-11.2"	
	1830..2326	
repeat_region	/note="LORa repeat: matches 1..497 of consensus"	
	3932..3965	
	/note="17 copies 2 mer gt 94% conserved"	
repeat_region	4632..5067	
	/note="CORb repeat: matches 1..461 of consensus"	
repeat_region	5733..5893	
	/note="LIMb3 repeat: matches 6002..6146 of consensus"	
	6000..6083	
	/note="MER4 repeat: matches 454..539 of consensus"	
repeat_region	6108..6171	
	/note="MER23 repeat: matches 993..1056 of consensus"	
repeat_region	7143..7840	
	/note="UTRb repeat: matches 1..691 of consensus"	
	8055..8292	
	/note="UTR16A repeat: matches 203..442 of consensus"	
repeat_region	8491..8686	
	/note="UTR29 repeat: matches 6..203 of consensus"	
repeat_region	8922..9049	
	/note="MERd repeat: matches 483..677 of consensus"	
repeat_region	9062..9281	
	/note="110 copies 2 mer tc 59% conserved"	
repeat_region	9398..9530	
	/note="MER1C repeat: matches 1..131 of consensus"	
repeat_region	9623..9732	
	/note="UTR1 repeat: matches 675..785 of consensus"	
repeat_region	9970..10333	
	/note="UTR2 repeat: matches 1..486 of consensus"	
repeat_region	10400..10545	
	/note="UTR9 repeat: matches 2..151 of consensus"	
repeat_region	10549..10935	
	/note="MER67D repeat: matches 12..391 of consensus"	
repeat_region	11249..11332	
	/note="MIR repeat: matches 69..144 of consensus"	
repeat_region	11923..12159	
	/note="MERd repeat: matches 1..234 of consensus"	
repeat_region	12164..12767	
	/note="MERd repeat: matches 396..973 of consensus"	
repeat_region	12852..13321	
	/note="THB1B repeat: matches 1..364 of consensus"	
repeat_region	13222..14798	
	/note="THB1B-INTERNAL repeat: matches 1..1580 of consensus"	
repeat_region	14799..15173	
	/note="THB1B repeat: matches 1..364 of consensus"	
repeat_region	16144..16346	
	/note="MER4B repeat: matches 585..789 of consensus"	
repeat_region	16339..16503	
	/note="MER4B repeat: matches 112..278 of consensus"	
repeat_region	16504..16571	
	/note="MER6A repeat: matches 1..69 of consensus"	
repeat_region	16570..16629	
	/note="MER1B repeat: matches 283..341 of consensus"	
repeat_region	16599..16902	
	/note="MER6A repeat: matches 163..478 of consensus"	
repeat_region	16903..17014	
	/note="MER4B repeat: matches 3..115 of consensus"	
repeat_region	17195..17491	
	/note="L1v1 repeat: matches 1..297 of consensus"	
repeat_region	18120..19291	
	/note="L1P repeat: matches 1726..2897 of consensus"	
repeat_region	19460..19584	
	/note="25 copies 5 mer attc 62% conserved"	
repeat_region	19686..19745	
	/note="12 copies 5 mer tttaa 71% conserved"	

```
repeat_region 20453..20914
/note="LIMD2 repeat: matches 5745..6199 of consensus"
repeat_region 20913..21222
/note="LIMD2 repeat: matches 5277..5594 of consensus"
repeat_region 21308..21570
/note="LTR9 repeat: matches 357..625 of consensus"
repeat_region 21645..21984
/note="LTR19B repeat: matches 55..393 of consensus"
repeat_region 22081..22583
/note="LIMD2 repeat: matches 4726..5242 of consensus"
repeat_region 23498..23784
/note="AluDb repeat: matches 1..305 of consensus"
repeat_region 23873..24169
/note="AluSg repeat: matches 1..295 of consensus"
repeat_region 25044..25238
/note="MIR repeat: matches 31..238 of consensus"
repeat_region 25404..25696
/note="AluSx repeat: matches 20..311 of consensus"
repeat_region 25832..26218
/note="MSTC repeat: matches 1..405 of consensus"
repeat_region 26221..27865
/note="MSTC-internal repeat: matches 2..1651 of consensus"
repeat_region 27907..28659
/note="LIPB2 repeat: matches 5400..6155 of consensus"
repeat_region 28710..29043
/note="MSTD repeat: matches 30..394 of consensus"
repeat_region 29634..29697
/note="32 copies 2 mer ta 79% conserved"
repeat_region 29729..29752
/note="12 copies 2 mer tg 95% conserved"
repeat_region 30688..30874
/note="MER1A repeat: matches 160..354 of consensus"
repeat_region 31648..31785
/note="LIMC4 repeat: matches 6409..6540 of consensus"
repeat_region 31813..32243
/note="LIMC4 repeat: matches 6629..7089 of consensus"
repeat_region 32449..32821
/note="THEIC repeat: matches 1..371 of consensus"
repeat_region 32824..34464
/note="THEIC-internal repeat: matches 3..1651 of consensus"
repeat_region 34465..34825
/note="THEIC repeat: matches 1..360 of consensus"
repeat_region 34804..35004
/note="MER67C repeat: matches 202..403 of consensus"
repeat_region 39792..39868
/note="LIPB12 repeat: matches 6076..6152 of consensus"
repeat_region 40953..41258
/note="AluY repeat: matches 1..308 of consensus"
repeat_region 41279..41308
/note="15 copies 2 mer aa {6% conserved"
repeat_region 42744..42933
/note="LIMB8 repeat: matches 6064..6256 of consensus"
repeat_region 42973..43275
/note="AluSx repeat: matches 1..303 of consensus"
repeat_region 43560..43991
/note="match: STS: Em:HSPH07E5"
repeat_region 45196..45584
/note="MIR2B repeat: matches 2..399 of consensus"
repeat_region 45594..45767
/note="87 copies 2 mer ta 77% conserved"
repeat_region 45800..45851
/note="MIR2B repeat: matches 394..448 of consensus"
repeat_region 48118..48487
/note="THEIC repeat: matches 1..370 of consensus"
repeat_region 48708..48806
/note="AluIo/FRAM repeat: matches 197..297 of consensus"
repeat_region 49853..50271
/note="LIMC repeat: matches 2274..2369 of consensus"
repeat_region 50637..50927
/note="LIME repeat: matches 974..1253 of consensus"
repeat_region 52250..52349
/note="MIR repeat: matches 131..228 of consensus"
```

```
repeat_region 52696..52996
/note="AluSg repeat: matches 1..293 of consensus"
repeat_region 54927..55010
Query Match 81.3% Score 407.2; DB 9; Length 140756;
Best Local Similarity 91.9%; Pred. No. 2,4e-120;
Matches 441; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 4 TCGGCAACCTCCCAACAGACCTTAGGTTTCCTGTGAGATGGGGACTGAGACAG 63
Db 69419 TTGGCAACTTCCCAACAGACTGGGTTTCTGTGAGAGGGGACTGAGACAG 69478
QY 64 GACTAGCTGATTTCCCTAGGCTGACTAAGATCCCTAAGCTTACTGGAGTGACAC 123
Db 69479 GACTAGTTGATTTCCCTAGGCACTAAGATCCCTAAGCTTACTGGAGTGACCG 69538
QY 124 ATCAACCTTTAAACAGGGGCTTCACTTAGCTCACACCTGACCAATCAGAGACTAC 183
Db 69539 ATCAACCTTTAAACAGGGGCTTCACTTAGCTCACACCTGACCAATCAGAGACTAC 69598
QY 184 TAAATGCTAATTAGCAAAAGACAGAGGTAAAGAAATAGCCAATCATATTGCTGAG 243
Db 69599 TAAATGCTAATTAGCAAAAGAGGTAAAGAAATAGCCAATCATATTGCTGAA 69658
QY 244 AGCAGCAGAGGAGACATATGCGGATATAAACCAAGTCTTGAGCCGGCAACGCA 303
Db 69659 AGCAGCAGGAGGAGACATATGCGGATATAAACCAAGTCTTGAGCCGGCAACGCA 69718
QY 304 ACCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCAATGCTACTAT 363
Db 69719 ACCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCAATGCTACTAT 69778
QY 364 TAAATCTGCACTGACTCTTCTGATGATGTTCTTAAAGGCTTGAAGCTGAGC-TTTCG 422
Db 69779 TAAATCTGCACTGACTCTTCTGATGATGTTCTTAAAGGCTTGAAGCTGAGCTTTTCG 69838
QY 423 CTCGCATCCACCACTGCTGTTTGGCCGACCGAGACCCCGCTGATCCCATCCCTC 482
Db 69839 CTCGCCTCCACCACTGCTGTTTGGCACTGTCGAGACCCCGCTGATCCCATCCCTC 69898
```

Search completed: April 19, 2003, 13:59:12
Job time : 2138.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:05:38 ; Search time 180.12 Seconds
(without alignments)
6263.892 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gctcgcgaacctccca.....ctgcatcagcaggtctcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10499	21	Human retroviral s
2	498	99.4	56093	24	Colo adenocarcino
3	461.4	92.1	7466	23	DNA encoding novel
4	450	89.8	711	21	Human retroviral s
5	426	85.0	1329	19	Multiple sclerosis
6	425.6	85.0	783	20	Clone 5M6 from MSR
7	396.2	79.1	711	21	Human endogenous r
8					Human retroviral s
9	396.2	79.1	1393	22	Human diagnostic a

10	377.8	75.4	893	23	AA65964	DNA encoding novel
11	377	75.2	2781	22	AA55630	Nucleotide sequenc
12	375.4	74.9	2946	20	AA77526	Human secreted pro
13	375.4	74.9	2946	21	AA259468	Human secreted pro
14	374.4	74.7	2930	24	AA24195	Human syncytin cDN
15	373.4	74.5	7582	20	AA25665	Complete human end
16	373.4	74.5	7582	21	AA59215	Human endogenous r
17	372.6	74.4	849	22	AA51000	Human diagnostic a
18	372.2	74.3	3372	20	AA25663	Human endogenous r
19	372.2	74.3	3372	21	AA59213	Partial pol gene a
20	371.2	74.1	1136	20	AA25660	Human endogenous r
21	371.2	74.1	1136	21	AA59210	Human endogenous r
22	371.2	74.1	3831	23	AA57127	3' pol gene and 3'
23	371.2	74.1	5154	23	AA567609	DNA encoding novel
24	371.2	74.1	8279	23	AA567474	DNA encoding novel
25	371.2	74.1	8294	23	AA584209	DNA encoding novel
26	370.6	74.0	2782	20	AA25661	Human endogenous r
27	370.6	74.0	2782	21	AA59211	5' non coding, 3'
28	370.6	74.0	2782	22	AA20069	HERV-W envelope pr
29	350.8	70.0	2300	23	AA592680	DNA encoding novel
30	347.4	69.3	1165	23	AA572721	DNA encoding novel
31	346	69.1	2942	23	AA577313	DNA encoding novel
32	345.8	69.0	2629	22	AA687491	Human immune/haema
33	341.8	68.2	635	19	AA43215	Multiple sclerosis
34	341.8	68.2	635	20	AA529702	Clone Cl6-3' from
35	341.8	68.2	2030	21	AA63826	Nucleotide sequenc
36	325.6	65.0	808	23	AA588392	DNA encoding novel
37	320.8	64.2	1478	23	AA587568	DNA encoding novel
38	320.8	64.0	1243	23	AA584189	DNA encoding novel
39	312.6	62.4	448	23	AA587567	DNA encoding novel
40	312.6	62.4	583	23	AA588391	DNA encoding novel
41	311.4	62.2	410	24	ABN94040	Gene #538 used to
42	311.4	61.5	17758	22	AA64210	Human benign prost
43	308.2	60.6	439	22	AA64138	Human immune/haema
44	303.4	60.6	439	22	AA63993	Human polynucleoti
45	303.4	60.6	439	22	AA51671	Genomic Sequence #

ALIGNMENTS

RESULT 1
ABN97929 standard; DNA; 10499 BP.
AC ABN97929;
DT 01-AUG-2002 (first entry)
XX
DE Human retroviral sequence HERV-7q.
XX
KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX
OS Human retrovirus.
XX
PN WO9967395-A1.
XX
PD 29-DEC-1999.
XX
PF 23-JUN-1999; 99WO-FR01513.
XX
PR 23-JUN-1998; 98FR-0007920.
XX
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Alliel PM, Perin J, Rieger F;
XX WPI; 2000-160587/14.
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
XX used for diagnosis, treatment and prevention of autoimmune and
XX neurological diseases -

XX Claim 3; Fig 1; 225pp; French.
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8.2e-161;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCTGGCCCACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGTGAAGA 60
Db 9500 GTCTGGCCCACTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGTGAAGA 9559
QY 61 CAGAGCTAGCTGAGATTTCTGAGCTGAGCACTAAGATCCCTAAGCTAGCTGGGAAGTGAC 120
Db 9560 CAGAGCTAGCTGAGATTTCTGAGCTGAGCACTAAGATCCCTAAGCTAGCTGGGAAGTGAC 9619
QY 121 CACATCCACCTTTAAACAAGGGGCTTGCACTAGTCAACACCTGACCACTAGAGAGCT 180
Db 9620 CACATCCACCTTTAAACAAGGGGCTTGCACTAGTCAACACCTGACCACTAGAGAGCT 9679
QY 181 CACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCATCTATTGCTT 240
Db 9680 CACTAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATGCCAATCATCTATTGCTT 9739
QY 241 GAGACACAGAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGCGCAACG 300
Db 9740 GAGACACAGAGGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCGCGCAACG 9799
QY 301 GCAACCCCTTTGGGCTCCCTCTTTGATGGAGTCTGTGTTTCACTGATTTTCACTC 360
Db 9800 GCAACCCCTTTGGGCTCCCTCTTTGATGGAGTCTGTGTTTCACTGATTTTCACTC 9859
QY 361 TATTAAATCTTGCAACTGACTCTTCTGTGCTCATGTTCTTAAAGGCTTGAGCTT 420
Db 9860 TATTAAATCTTGCAACTGACTCTTCTGTGCTCATGTTCTTAAAGGCTTGAGCTT 9919
QY 421 CGCTGCGCATCCACAGACGCTGTTGCGCGCACCGCAACCGCGCGCTGACCTCCATCC 480
Db 9920 CGCTGCGCATCCACAGACGCTGTTGCGCGCACCGCAACCGCGCGCTGACCTCCATCC 9979
QY 481 TCTGGATCATGAGGGGTGTC 501
Db 9980 TCTGGATCATGAGGGGTGTC 10000

KM gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-231133P.
PR 18-SEP-2000; 2000US-231617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237178P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (i)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL6164
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (i) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
SQ Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;
Query Match 99.4%; Score 498; DB 24; Length 56093;
Best Local Similarity 100.0%; Pred. No. 2,2e-159;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 37504 TCGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 37563
QY 64 GACTAGCTGATTTCTTGGCTGATTAAGATCCCTAAGCCTGAGTGGAAAGTACAC 123
DB 37564 GACTAGCTGATTTCTTGGCTGATTAAGATCCCTAAGCCTGAGTGGAAAGTACAC 37623
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 183
DB 37624 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 37683
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATAGCCAAATCATCTTATTTGCTGAG 243
DB 37684 TAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATAGCCAAATCATCTTATTTGCTGAG 37743
QY 244 AGCAGAGAGAGAGAGCAATGATCGGGATATTAACCCAGTCTTGAGAGCCGGCAACGGCA 303
DB 37744 AGCAGAGAGAGAGAGCAATGATCGGGATATTAACCCAGTCTTGAGAGCCGGCAACGGCA 37803
QY 304 ACCCCCTTTGGGTCCTCCCTCTTTGATGGAGAGCTGTGTTTCATGCTATTTCACCTAT 363
DB 37804 ACCCCCTTTGGGTCCTCCCTCTTTGATGGAGAGCTGTGTTTCATGCTATTTCACCTAT 37863
QY 364 TAAATCTTGCACTGACCTCTTCTGTCATGTTTCTTAACGGCTTGAGAGCTTTCGCG 423
DB 37864 TAAATCTTGCACTGACCTCTTCTGTCATGTTTCTTAACGGCTTGAGAGCTTTCGCG 37923
QY 424 TCGCCATCCACCTGCTGTTTGGCCGACCGGACACCGCGCTGACTCCCATCCCTCT 483
DB 37924 TCGCCATCCACCTGCTGTTTGGCCGACCGGACACCGCGCTGACTCCCATCCCTCT 37983
QY 484 GGATCATGAGGGGTGCC 501
DB 37984 GGATCATGAGGGGTGCC 38001
RESULT 3
AAS68626
ID AAS68626 standard; cDNA, 7466 BP.
XX
AC AAS68626;
XX
DT 13-FEB-2002 (first entry)
XX
DB DNA encoding novel human diagnostic protein #4430.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX OS Homo sapiens.
XX FN WO200175067-A2.
XX PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2001; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSEQ-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG04439.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 4430; 103pp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from Wipro
CC at ftp.wipro.int/pub/published_pct_sequences.
XX
SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;
Query Match 92.1%; Score 461.4; DB 23; Length 7466;
Best Local Similarity 96.6%; Pred. No. 2.8e-147;
Matches 482; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
QY 4 TCGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 3668 TCGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 3727
QY 64 GACTAGCTGATTTCTTGGCTGATTAAGATCCCTAAGCCTGAGTGGAAAGTACAC 123
DB 3728 GACTAGCTGATTTCTTGGCTGATTAAGATCCCTAAGCCTGAGTGGAAAGTACAC 3787
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 183
DB 3788 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 3847
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATAGCCAAATCATCTTATTTGCTGAG 243
DB 3848 TAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATAGCCAAATCATCTTATTTGCTGAG 3907
QY 244 AGCAGAGAGAGAGAGCAATGATCGGGATATTAACCCAGTCTTGAGAGCCGGCAACGGCA 303
DB 3908 AGCAGAGAGAGAGAGCAATGATCGGGATATTAACCCAGTCTTGAGAGCCGGCAACGGCA 3967
QY 304 -ACCCCTTTGGGTCCTCCCTCTTTGATGGAGAGCTGTTTCATGCTATTTCACCTCA 362
DB 3968 GCCCCTTTGGGTCCTCCCTCTTTGATGGAGAGCTGTTTCATGCTATTTCACCTCA 4027
QY 363 TAAATCTTGCACTGACCTCTTCTGTCATGTTTCTTAACGGCTTGAGAGCTTTCG 422

Db 4028 TTAATCTTGCACTGACTCTTCTGATCJATGTTTCTTAAGCGCTGAGCTTGTG 4087
 QY 423 CTGCGCATCCACCACTGCTGTTTGCGGCAACCGAGACCCGGCGTGACTCCCATCCCTC 482
 Db 4088 CTCACCGTCCACCACTGCTGTTTGCCACCAACCGAGACCTGCGCTGACTCCCATCCCTC 4147
 QY 483 TGGATCATGCAAGGTGTCTC 501
 Db 4148 TGGATCTGTCAGAGGTGTCTC 4166

RESULT 4

ABN97947
 ID ABN97947 standard; DNA; 711 BP.

XX AC ABN97947;

XX DT 01-AUG-2002 (first entry)

XX DE Human retroviral sequence R1F.

XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
 XX KM multiple sclerosis; ds.

XX OS Human retrovirus.

XX PN MO9967395-A1.

XX PD 29-DEC-1999.

XX PF 23-JUN-1999; 99WO-FR01513.

XX PR 23-JUN-1998; 98FR-0007920.

XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX PI Allele PM, Perin J, Rieger F;

XX DR WPI; 2000-160587/14.

XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 XX PS used for diagnosis, treatment and prevention of autoimmune and
 XX PT neurological diseases -

XX PS Claim 3; Fig 3; 225pp; French.

XX CC The present invention relates to new nucleic acid sequences of human
 XX CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 XX CC Regulatory elements associated with HERV-7q may alter expression of other
 XX CC genes (even remote genes) on the same chromosome, inducing immunological
 XX CC and/or neurological changes (which may be pathological or protective/
 XX CC curative). HERV-7q peptides can be used to improve efficiency of the
 XX CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 XX CC sequences can be used in immunogenic or vaccinating compositions, for
 XX CC protection against autoimmune diseases, particularly multiple sclerosis.
 XX CC The peptides may also be used (by sequence comparison) to detect/identify
 XX CC endogenous retroviruses that are abnormally expressed in cancer.
 XX CC neuropeptides or other autoimmune diseases. The present sequence was
 XX CC used to illustrate the invention.

XX CC Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;

XX CC Query Match 89.8%; Score 450; D3 21; Length 711;
 XX CC Best Local Similarity 100.0%; Pred. No. 6; 4e-144;
 XX CC Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 ACTGAGAGACAGACTAGCTGATTCTTCTAGGCTGACTAAGATCCCTAAGCTTACTG 111
 Db 1 ACTGAGAGACAGACTAGCTGATTCTTCTAGGCTGACTAAGATCCCTAAGCTTACTG 60

QY 112 GAAGGTGACCACTCCACTTTTAAACAGCGGGCTTGCACTTAGCTCACAAGCTGACCAAT 171
 Db 61 GAAGGTGACCACTCCACTTTTAAACAGCGGGCTTGCACTTAGCTCACAAGCTGACCAAT 120

QY 172 CAGAGAGCTCACTAAATGCTAATTTAGCAAGA CAGAGGTAAAGAAATAGCCAATCAT 231
 Db 121 CAGAGAGCTCACTAAATGCTAATTTAGCAAGA CAGAGGTAAAGAAATAGCCAATCAT 180
 QY 232 CTATTGCTTGGAGACACAGCAGAGGGACATGATCGGATATTAACCAAGCTTTGAG 291
 Db 181 CTATTGCTTGGAGACACAGCAGAGGGACATGATCGGATATTAACCAAGCTTTGAG 240
 QY 292 CCGGCAACGCAACCCCTTTGGGTGCCCTCTTGTATGGAGCTCTGTTTCATGCT 351
 Db 241 CCGGCAACGCAACCCCTTTGGGTGCCCTCTTGTATGGAGCTCTGTTTCATGCT 300
 QY 352 ATTTCACCTATTAATTTTCAACCTGACACTTCTGCTCCATGTTTCTTAAGCTTGA 411
 Db 301 ATTTCACCTATTAATTTTCAACCTGACACTTCTGCTCCATGTTTCTTAAGCTTGA 360
 QY 412 CTGAGCTTTGGCTGGCCATCCACCACTGCTGTTTGGCCGACCCGACAGCCGGCTGAC 471
 Db 361 CTGAGCTTTGGCTGGCCATCCACCACTGCTGTTTGGCCGACCCGACAGCCGGCTGAC 420
 QY 472 TCCCATCCCTCGATCATGCAAGGTGTCTC 501
 Db 421 TCCCATCCCTCGATCATGCAAGGTGTCTC 450

RESULT 5

AAV43219
 ID AAV43219 standard; CDNA; 1329 BP.

XX AC AAV43219;

XX DT 29-DEC-1998 (first entry)

XX DE Multiple sclerosis associated retrovirus fragment 6.

XX KW Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
 XX KW gag gene; env gene; rheumatoid arthritis-associated virus; ss.

XX OS Multiple sclerosis associated retrovirus.

XX PN MO9823755-A1.

XX PD 04-JUN-1998.

XX PF 26-NOV-1997; 97WO-IB01482.

XX PR 26-NOV-1996; 96US-0756429.

XX PA (INMR) BIO MERIEUX.

XX PI Bedin F, Beeseme F, Jolivet-Reynaud C, Komurian-Pradel F;
 XX PI Mandrand B, Paranhos-Baccala G, Perron H;

XX DR WPI; 1998-322732/28.

XX PF P-PSDB; AAW71069.

XX PT New nucleic acid from retroviruses - useful for diagnosis,
 XX PT prevention and treatment of, e.g. multiple sclerosis

XX PS Disclosure; Pages 187-188; 286pp; English.

CC The present sequence represents a multiple sclerosis (MS) associated
 CC retrovirus (MSRV) genomic fragment used in the method of the
 CC invention. The invention provides complete or partial genomic
 CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
 CC polypeptides encoded by these genes. The invention also provides
 CC antibodies raised against the polypeptides. The genomic sequences,
 CC polypeptides and antibodies are also claimed useful for diagnosing
 CC infection by MS and rheumatoid arthritis-associated viruses, and also
 CC for prevention and treatment of infection with these viruses.

XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 19; Length 1329;

Best Local Similarity 90.6%; Pred. No. 1.6e-135; Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGGCCAAGCTCCCAAGACACTTGTGTTTCTGTTGAGAGGGGAGTGAAGAC 61
 Db TGTGAGCCAACTCCCAAGACACTTGTGTTTCTGTTGAGAGGGGAGTGAAGAC 556
 QY 62 AGAAGTACCTGATTTCTAGAGCTGACTAAGATCCCTAAGGCTAGTGGAGAGTGAAC 121
 Db 557 AGGACTAGCTGATTTCTAGAGCTGACTAAGATCCCTAAGGCTAGTGGAGAGTGAAC 616
 QY 122 ACATCCACCTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGAGCTC 181
 Db 617 GCATCATCTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGAGCTC 676
 QY 182 ACTAAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCATCATCTATTGCTTG 241
 Db 677 ACTAAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCATCATCTATTGCTTG 736
 QY 242 AGAGCAGCAGAGAGAGGAGCAATGATCGGATATTAACCCAAATCTTCGAGCGGCAAG 301
 Db 737 AGAGCAGCAGAGAGAGGAGCAATGATCGGATATTAACCCAAATCTTCGAGCGGCAAG 796
 QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTGTTTTCACTATTTCACTCT 361
 Db 797 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTGTTTTCACTATTTCACTCT 856
 QY 362 ATTAATCTTGAACCTGACTTTCTGTTGATGTTTCTTACGCTTGAAGTGAAGCTTTC 421
 Db 857 ATTAATCTTGAACCTGACTTTCTGTTGATGTTTCTTACGCTTGAAGTGAAGCTTTC 916
 QY 422 GGTGGCCATCCACCACTGCTGTTTGGCCGACCGGCGGCTGATCCCATCCCT 481
 Db 917 GTTGGCCATCCACCACTGCTGTTTGGCCGACCGGCGGCTGATCCCATCCCT 976
 QY 482 CTGATCATGACGAGGTGTC 501
 Db 977 TTGATCCAGCAGAGTGTCC 996

RESULT 6
 AAX29704
 ID AAX29704 standard; DNA; 1329 BP.

XX AAX29704;
 AC 08-JUN-1999 (first entry)
 DT Clone 5M6 from MSRV-1.
 DE Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 KM rheumatoid polyarthritits; ss.
 XX Multiple sclerosis related virus type 1.
 OS FR2765588-A1.
 PN 08-JAN-1999.
 PD 07-JUL-1997; 97FR-0008816.
 XX PF

XX 07-JUL-1997; 97FR-0008816.
 PR (INMR) BIO MERIEUX.
 XX WPI; 1999-098275/09.
 DR P-PSDB; AAW99554.
 XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
 PT multiple sclerosis or rheumatoid polyarthritits
 PS Claim 1; Page 39-40; 83pp; French.
 XX This sequence represents clone 5M6 from a novel multiple sclerosis
 CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
 CC prophylactic or therapeutic compositions to inhibit expression of a
 CC multiple sclerosis related virus and/or virus associated with
 CC rheumatoid polyarthritits.

XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 20; Length 1329;

Best Local Similarity 90.6%; Pred. No. 1.6e-135; Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGGCCAAGCTCCCAAGACACTTGTGTTTCTGTTGAGAGGGGAGTGAAGAC 61
 Db 497 TGTGAGCCAACTCCCAAGACACTTGTGTTTCTGTTGAGAGGGGAGTGAAGAC 556
 QY 62 AGAAGTACCTGATTTCTAGAGCTGACTAAGATCCCTAAGGCTAGTGGAGAGTGAAC 121
 Db 557 AGGACTAGCTGATTTCTAGAGCTGACTAAGATCCCTAAGGCTAGTGGAGAGTGAAC 616
 QY 122 ACATCCACCTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGAGCTC 181
 Db 617 GCATCATCTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGAGCTC 676
 QY 182 ACTAAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCATCATCTATTGCTTG 241
 Db 677 ACTAAATGCTAATTAAGCAAGAGAGTAAAGAAATAGCCATCATCTATTGCTTG 736
 QY 242 AGAGCAGCAGAGAGAGGAGCAATGATCGGATATTAACCCAAATCTTCGAGCGGCAAG 301
 Db 737 AGAGCAGCAGAGAGAGGAGCAATGATCGGATATTAACCCAAATCTTCGAGCGGCAAG 796
 QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTGTTTTCACTATTTCACTCT 361
 Db 797 CAACCCCTTTGGGTCCCTCCCTTTGATAGGAGCTGTTTTCACTATTTCACTCT 856
 QY 362 ATTAATCTTGAACCTGACTTTCTGTTGATGTTTCTTACGCTTGAAGTGAAGCTTTC 421
 Db 857 ATTAATCTTGAACCTGACTTTCTGTTGATGTTTCTTACGCTTGAAGTGAAGCTTTC 916
 QY 422 GGTGGCCATCCACCACTGCTGTTTGGCCGACCGGCGGCTGATCCCATCCCT 481
 Db 917 GTTGGCCATCCACCACTGCTGTTTGGCCGACCGGCGGCTGATCCCATCCCT 976
 QY 482 CTGATCATGACGAGGTGTC 501
 Db 977 TTGATCCAGCAGAGTGTCC 996

RESULT 7
 AAX25669
 ID AAX25669 standard; cDNA to mRNA; 783 BP.

XX AAX25669;
 AC 21-MAY-1999 (first entry)
 DT Human endogenous retrovirus w long terminal repeat region.
 DE Clone; human endogenous retrovirus; genome; autoimmune disease;
 XX KW

KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

OS Human endogenous retrovirus.

XX WO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INRM) BIO MERIEUX.

XX Beeseme F, Blond JL, Boucon O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 83; 106pp; French.

XX This sequence represents the long terminal repeat (LTR) region of the
 CC human endogenous retrovirus (HERV) W genome. The nucleic acids, their
 CC fragments or peptides encoded by them are markers of autoimmune disease
 CC (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus
 CC erythematosus, insulin-dependent diabetes and related pathologies) and
 CC of abnormal or unsuccessful pregnancy and can be used as chromosomal
 CC markers for susceptibility to these conditions, or proximity markers
 CC of genes associated with this susceptibility.

XX Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;

XX Query Match 85.0%; Score 425.6; DB 20; Length 783;

XX Best Local Similarity 94.0%; Pred. No. 1.6e-135;

XX Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;

QY 54 TTAGAGACGAGCTAGCTGATTTCTTAGCTGACCTAGAAATCCCTTAAGCTAGTGGGA 113

DB 1 TTAGAGACGAGCTAGCTGATTTCTTAGCTGACCTAGAAATCCCTTAAGCTAGTGGGA 60

QY 114 AGGTGACCATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCAGACCTGACCAATCA 173

DB 61 AGGTGACCATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCAGACCTGACCAATCA 120

QY 174 GAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTTAAAGAAATAGCCATCATCT 233

DB 121 GAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTTAAAGAAATAGCCATCATCT 180

QY 234 ATTGCTGAGACACGAGGAGGACATGATCCGGATTTAAACCAAGCTCTGAGGC 293

DB 181 ATTGCTGAGACACGAGGAGGACATGATCCGGATTTAAACCAAGCTCTGAGGC 240

QY 294 GGCACAGGCA-AACCCCTTTGGTCCCTCCCTTTGTATGGAGCTCTTTTCAATGCTA 352

DB 241 GGCACAGGCA-AACCCCTTTGGTCCCTCCCTTTGTATGGAGCTCTTTTCAATGCTA 300

QY 353 TTTCACCTATTAAATCTTGCAACGTCACTCTTGCTGTCATGTTTCTTAAGGCTTGAAC 412

DB 301 TTTCACCTATTAAATCTTGCAACGTCACTCTTGCTGTCATGTTTCTTAAGGCTTGAAC 360

QY 413 TTAGCTTTGGCTGCGCATCCACATGCTGTTTGGCCGACCGGACAGCCGGCGCTGACT 472

DB 361 TTAGCTTTGGCTGCGCATCCACATGCTGTTTGGCCGACCGGACAGCCGGCGCTGACT 420

QY 473 CCATCCCTCTGGATCATGAGGGTGTCC 501

DB 421 CCATCCCTCTGGATCATGAGGGTGTCC 449

RESULT 8

ABN97946

ABN97946 standard; DNA; 711 BP.

AC ABN97946;

DT 01-AUG-2002 (first entry)

XX Human retroviral sequence R1.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX multiple sclerosis; ds.

XX Human retrovirus.

XX WO9967395-A1.

XX 29-DEC-1999.

XX 23-JUN-1999; 99WO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,

XX used for diagnosis, treatment and prevention of autoimmune and

XX neurological diseases

XX Claim 3; Fig 3; 225pp; French.

XX The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

XX Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;

XX Query Match 79.1%; Score 396.2; DB 21; Length 711;

XX Best Local Similarity 95.3%; Pred. No. 2e-125;

XX Matches 430; Conservative 0; Mismatches 18; Indels 3; Gaps 2;

QY 52 ACTGAGACAGGACTAGCTGATTTCTTAGCTGACCTAAGAAATCCCTAAGCTAGCTGG 111

DB 1 ACTGAGACAGGACTAGCTGATTTCTTAGCTGACCTAAGAAATCCCTAAGCTAGCTGG 60

QY 112 GAAAGTGACACATCCACTTTAAACAGGGGCTTGCACCTAGCTCACACTGACCAAT 171

DB 61 GAAAGTGACACATCCACTTTAAACAGGGGCTTGCACCTAGCTCACACTGACCAAT 120

QY 172 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTTAAAGAAATAGCCATCAT 231

DB 121 CAGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTTAAAGAAATAGCCATCAT 180

QY 232 CTATGCTGAGAGACAGAGAGGACCAATGATGGGATTTAAACCAAGCTTGGAG 291

DB 181 CTATGCTGAGAGACAGAGAGGACCAATGATGGGATTTAAACCAAGCTTGGAG 240

QY 292 CCGGCAAGGCA-AACCCCTTTGGTCCCTCCCTTTGTATGGAGCTCTTTTCAATGC 350

|||||
Db 547 TCACCTATTAATCTTCACTGACCTCTCTGTGTGTTGTTGAGCTG 606
Qy 415 AGCTTGGCTGCCATTCACACACTGCTTTGGCCGACCCGAGACCGCGCTGACTCC 474
Db 607 AGCTTGGCTGCCATTCACACACTGCTTTGGCCGACCCGAGACCTGCGCTGACTTC 666
Qy 475 CATCCCTCTGATCATGACGAGGTGTCC 501
Db 667 CATCCCTCTGATCATGACGAGGTGTTC 693
RESULT 10
AAS65964
ID AAS65964 standard; cDNA; 893 BP.
XX
AC AAS65964;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #1768.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PP 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HXSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR P-PSDB; AAG01777.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 1768; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;
SQ

Query Match 75.4%; Score 377.8; DB 23; Length 893;
Best Local Similarity 88.6%; Pred. No. 4,7e-119;
Matches 451; Conservative 0; Mismatches 37; Indels 21; Gaps 3;
Qy 4 TCGGCCAACCTCCCCACACACACTAGTCTTCTCTGTAGATGGGGGACTGAGACAG 63
Db 188 TCGGCCAACCTCCCCACACACACTAGTCTTCTCTGTAGATGGGGGACTGAGACAG 247
Qy 64 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTTAGCTAGCTGGAGGTACAC 123
Db 248 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTTAGCTAGCTGGAGGTACAC 307
Qy 124 ATCCACCTTTAACAAGGGGCTTGCACCTTAGCTACACCTGACCATC-----AG 174
Db 308 ATCCACCTTTAACAAGGGGCTTGCACCTTAGCTACACCTGACCATCAGTAAGAAAG 367
Qy 175 AGAGCTACCTAAATGCTAATTAGCAAAAGCAGGAGGTAAAGAAATAGCAATCATCTA 234
Db 368 AGAGCTACCTAAATGCTAATTAGCAAAAGCAGGAGGTAAAGAAATAGCAATCATCTA 427
Qy 235 TTGCTTGAAGACACAGAGGAGGACATGATCGGATATTAACCAAGTCTTGAGCCG 294
Db 428 TTGCTTGAAGACACAGGAGGAGGACATGATCGGATATTAACCAAGGCTTGAGCCG 487
Qy 295 GCAACGGCAACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTGTTTCATGCTATT 354
Db 488 GCAACGGCAACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTGTTTCATGCTATT 537
Qy 355 TCACCTATTAAATCTTGCAACCTG--CACTCTGAGGACATGTTTCTTAGGCTTGAGC 412
Db 538 TCACCTATTAAATCTTGCAACCTG--CACTCTGAGGACATGTTTCTTAGGCTTGAGC 597
Qy 413 TGAGCTTTGCTCGCCATCCACCACTGCTGTTTGCAGCCGACCGAGCCGCTGACT 472
Db 598 TGAGCTTTGCTCGCCATCCACCACTGCTGTTTGCAGCCGAGCCGCTGACT 657
Qy 473 CCATTCCTCTGATCATGACGAGGTGTCC 501
Db 658 TCACCCCTCCAGATATGCGAGGGGTGTCC 686
RESULT 11
AAS55630
ID AAS55630 standard; DNA; 2781 BP.
XX
AC AAS55630;
XX
DT 29-MAY-2001 (first entry)
XX
XX Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
XX Envelope protein; HERV; syncytia formation; placental development;
XX syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
XX Key Location/Qualifiers
XX CDS 762..2378
XX FT /*tag= a
XX FT /product= "envelope protein"
XX
XX WO200116171-A1.
XX
XX PD 08-MAR-2001.
XX
XX PF 01-SEP-2000; 2000WO-FR02429.
XX
XX PR 01-SEP-1999; 99FR-0011141.
XX PR 15-SEP-1999; 99FR-0011793.
XX
XX (INRM) BIO MERIEUX
XX (INRM) INST NAT SANTE & RECH MEDICALE.

XX Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
PI WPI: 2001-22676/23.
DR P-PSDB; AAB67652.
XX
PT Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia -
XX
PS Disclosure; Page 44-45; 57pp; French.
XX
CC The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV), in
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.
XX
SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;
XX
Query Match 75.2%; Score 377; DB 22; Length 2781;
Best Local Similarity 100.0%; Pred. No. 1.6e-118; Mismatches 0; Gaps 0;
Matches 377; Conservative 0; Indels 0; Gaps 0;
XX
QY 4 TCGGCAACCTCCCAACAGCATTAGGTTTCTGTGAGATGGGGACTGAGAGACG 63
Db 2387 TCGGCAACCTCCCAACAGCATTAGGTTTCTGTGAGATGGGGACTGAGAGACG 2446
XX
QY 64 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCTGAGGAGTGACAC 123
Db 2447 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCTGAGGAGTGACAC 2506
XX
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTAGCTCACACCTGACCAATCAGAGACTCAC 183
Db 2507 ATCCACCTTTAAACAGGGGGCTTGCACTAGCTCACACCTGACCAATCAGAGACTCAC 2566
XX
QY 184 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
Db 2567 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 2626
XX
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 303
Db 2627 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 2686
XX
QY 304 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTGTGTTTCAATGCTATTCTCTAT 363
Db 2687 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTGTGTTTCAATGCTATTCTCTAT 2746
XX
QY 364 TAAATCTTGCAACTGCA 380
Db 2747 TAAATCTTGCAACTGCA 2763
XX
RESULT 12
AAK77526
ID AAK77526 standard; cDNA; 2946 BP.
XX
AC AAK77526;
XX
DT 10-AUG-1999 (first entry)
DE Human secreted protein AJ172_2 cDNA.
XX
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;

KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
OS Homo sapiens.
XX W09926972-A1.
XX
PD 03-JUN-1999.
XX
PF 17-NOV-1998; 98WO-US24614.
XX
PR 20-OCT-1998; 98US-0175928.
PR 21-NOV-1997; 97US-0976110.
PR 18-MAY-1998; 98US-0080478.
XX
PA (GEMV) GENETICS INST INC.
XX
FI Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;
PI McCoy JM, Merberg D, Treacy M;
XX
DR WPI: 1999-357813/30.
DR P-PSDB; AA08622.
XX
XX New polynucleotides encoding secreted proteins
PT
PS Claim 13a; Page 100-101; 142pp; English.
XX
CC This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful
CC for gene therapy.
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;
XX
Query Match 74.9%; Score 375.4; DB 20; Length 2946;
Best Local Similarity 99.7%; Pred. No. 6e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 4 TCGGCAACCTCCCAACAGCATTAGGTTTCTGTGAGATGGGGACTGAGAGACG 63
Db 2553 TCGGCAACCTCCCAACAGCATTAGGTTTCTGTGAGATGGGGACTGAGAGACG 2612
XX
QY 64 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCTGAGGAGTGACAC 123
Db 2613 GACTAGCTGATTTCTCTAGGCTGATAGATCCCTAAGCTGAGGAGTGACAC 2672
XX
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTAGCTCACACCTGACCAATCAGAGACTCAC 183
Db 2673 ATCCACCTTTAAACAGGGGGCTTGCACTAGCTCACACCTGACCAATCAGAGACTCAC 2732
XX
QY 184 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
Db 2733 TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 2792
XX
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 303
Db 2793 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 2852
XX
QY 304 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTGTGTTTCAATGCTATTCTCTAT 363
Db 2853 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTGTGTTTCAATGCTATTCTCTAT 2912

Qy 364 TAAATCTTGCAACTGCA 380
Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 13
AAZ59468 standard; cDNA; 2946 BP.
AAZ59468;
AAZ59468;
11-APR-2000 (first entry)
Human secreted protein Aji172_2 polynucleotide sequence.

Human: secreted protein; disease diagnosis; pre-eclampsia; cancer; placental pathology; metastasis inhibitor; nutritional activity; immune stimulator; haematopoiesis regulator; tissue growth; tumour inhibitor; anti-inflammatory; clove Aji172_2; ATCC_98115; gene therapy; ss.

Homo sapiens.
MO9960020-A1.
25-NOV-1999.
17-MAY-1999; 99WO-US10915.
18-MAY-1998; 98US-0080478.
20-OCT-1998; 98US-0175928.
(GEMV) GENETICS INST INC.
Jacob K, McCoy JM, Lavalie ER, Collins-Racie LA, Evans C; Mebergy D, Mi S, Treacy M; WPI; 2000-116311/10.
P-PSDB; AAY67313.

New polynucleotides encoding secreted cDNA libraries, used to develop products for the diagnosis and treatment of neoplastic disease -

Claim 14; Page 107-108; 1499p; English.

This is the human secreted protein Aji172_2 nucleotide sequence, obtained from a human adult testes cDNA library. The invention relates to secreted human and murine proteins. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Detection of the levels of the proteins can be used for the diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate the expression or function of the proteins may be used for treating a neoplastic disease and inhibiting metastasis. Other suggested activities include nutritional activity (e.g. in feeds), cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, haematopoiesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide sequences are also stated to be useful for gene therapy.

Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

Query Match 74.9%; Score 375.4; DB 21; Length 2946;
Best Local Similarity 99.7%; Pred. No. 66-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TGGGCAACTCCCAAGCACTTGGCTTCTTGAATGGGGAGCTAGAGACG 63
Db 2553 TGGGCAACTCCCAAGCACTTGGCTTCTTGAATGGGGAGCTAGAGACG 2612

Qy 64 GACTAGCTGATTTCTTGAAGCTGACTAAGAAATCCCTTAAGCTTAGCTGGGAAGTGACAC 123
Db 2613 GACTAGCTGATTTCTTGAAGCTGACTAAGAAATCCCTTAAGCTTAGCTGGGAAGTGACAC 2672

Qy 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATTCAGAGCTCAC 183
Db 2673 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAACCTGACCAATTCAGAGCTCAC 2732

Qy 184 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCAATCATATTGCTGAG 243
Db 2733 TAAATGCTAATTAGGCAAGAAACAGAGGTAAAGAAATAGCAATCATATTGCTGAG 2792

Qy 244 AGCAGAGAGAGGAGCAATGATGGATATAAACCAAGCTTCGAGCCGGCAAGCGCA 303
Db 2793 AGCAGAGAGAGGAGCAATGATGGATATAAACCAAGCTTCGAGCCGGCAAGCGCA 2852

Qy 304 ACCCCCTTTGGGGTCCCTCCCTTTGTATGGAGGCTGTTTTCATGCTATTCACTCAT 363
Db 2853 ACCCCCTTTGGGGTCCCTCCCTTTGTATGGAGGCTGTTTTCATGCTATTCACTCAT 2912

Qy 364 TAAATCTTGCAACTGCA 380
Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 14
AAD24195 standard; cDNA; 2930 BP.
AAD24195;
07-MAY-2002 (first entry)
Human syncytin cDNA.

Human: syncytin; pre-eclampsia; gestational trophoblast disorder; choriocarcinoma; hydatidiform mole; placental site tumour; abortion; envelope gene; human endogenous defective retrovirus; HERV-W; ss.

Homo sapiens.

Key Location/Qualifiers
CDS 930..2546
FT /*tag= a
FT /product= "Syncytin"

MO200204678-A2.
17-JAN-2002.
09-JUL-2001; 2001WO-US21719.
07-JUL-2000; 2000US-216657P.
(GEMV) GENETICS INST INC.
Keith JC, McCoy JM, Mi S; WPI; 2002-117127/22.
P-PSDB; AAE1540.

Identifying a compound for treating a subject with or at risk of developing pre-eclampsia, comprises determining whether the expression of activity of syncytin in the cell is modulated in the presence of a test compound -

Disclosure; Page 39-42; 43p; English.

The invention relates to identifying compounds which are modulators of syncytin expression. The syncytin modulators are useful in diagnosis and treatment of pre-eclampsia and gestational trophoblast disorders (e.g. choriocarcinoma, hydatidiform mole, placental site tumour and missed/incomplete abortion). Syncytin is a human gene derived from the

CC envelope gene of human endogenous defective retrovirus, HERV-W. The
CC present invention is based partly on the discovery that syncytin
CC expression is dramatically reduced in preclampsia, and is also
CC mis-localised to the apical syncytiotrophoblast membrane. The present
CC sequence is human syncytin cDNA.

XX Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;

Query Match 74.7%; Score 374.4; DB 24; Length 2930;

Best Local Similarity 99.7%; Pred. No. 1.3e-117; Indels 0; Gaps 0;

Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63

Db 2555 TCGGCCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 2614

QY 64 GACTAGCTGATTTCTTAGGCTGATTAAGATCCCTAAGCTAGTGGGAAGTACAC 123

Db 2615 GACTAGCTGATTTCTTAGGCTGATTAAGATCCCTAAGCTAGTGGGAAGTACAC 2674

QY 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 183

Db 2675 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 2734

QY 184 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCTATTGCTGAG 243

Db 2735 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCTATTGCTGAG 2794

QY 244 AGCAGAGAGAGGAGCAATGATCGGGATTAACCCCAAGCTTTGAGCCGGCAACGGCA 303

Db 2795 AGCAGAGAGAGGAGCAATGATCGGGATTAACCCCAAGCTTTGAGCCGGCAACGGCA 2854

QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCATGCTATTCTAT 363

Db 2855 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCATGCTATTCTAT 2914

QY 364 TAAATCTTGCAACTGC 379

Db 2915 TAAATCTTGCAACTGC 2930

RESULT 15

AAAX25665 standard; cDNA to mRNA; 7582 BP.

XX AAX25665;

XX 21-MAY-1999 (first entry)

XX Complete human endogenous retrovirus W genome.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;

XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

XX Human endogenous retrovirus.

XX WO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond JL, Boucon O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -

XX expressed exclusively in placenta and useful in diagnosis and

PT therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 71-74; 106pp; French.

XX This sequence represents the complete sequence of the human endogenous

XX retrovirus (HERV) W genome. The nucleic acids, their fragments or

XX peptides encoded by them are markers of autoimmune disease (e.g. multiple

XX sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,

XX insulin-dependent diabetes and related pathologies) and of abnormal or

XX unsuccessful pregnancy and can be used as chromosomal markers for

XX susceptibility to these conditions, or proximity markers of genes

XX associated with this susceptibility.

XX Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 74.5%; Score 373.4; DB 20; Length 7582;

Best Local Similarity 97.6%; Pred. No. 4.9e-117; Indels 0; Gaps 0;

Matches 368; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63

Db 7206 TCGGCCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 7265

QY 64 GACTAGCTGATTTCTTAGGCTGATTAAGATCCCTAAGCTAGTGGGAAGTACAC 123

Db 7266 GACTAGCTGATTTCTTAGGCTGATTAAGATCCCTAAGCTAGTGGGAAGTACAC 7325

QY 124 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 183

Db 7326 ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 7385

QY 184 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCTATTGCTGAG 243

Db 7386 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCTATTGCTGAG 7445

QY 244 AGCAGAGAGAGGAGCAATGATCGGGATTAACCCCAAGCTTTGAGCCGGCAACGGCA 303

Db 7446 AGCAGAGAGAGGAGCAATGATCGGGATTAACCCCAAGCTTTGAGCCGGCAACGGCA 7505

QY 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCATGCTATTCTAT 363

Db 7506 ACCCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGTGTTTCATGCTATTCTAT 7565

QY 364 TAAATCTTGCAACTGCA 380

Db 7566 TAAATCTTGCAACTGCR 7582

Search completed: April 19, 2003, 12:18:34
Job time : 205.12 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: April 19, 2003, 12:03:58 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000
Perfect score: 1 gttctggcaccactcccca.....ctgatcatgaggggtgtcc 501
Sequence:

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	375.4	74.9	2946	US-09-175-928-3	Sequence 3, Appli
2	238.4	47.6	279	US-08-686-878A-50	Sequence 50, Appli
3	238.4	47.6	279	US-08-721-489-4	Sequence 4, Appli
4	46.8	9.3	7218	US-08-233-463-14	Sequence 14, Appli
5	34.2	6.8	80246	US-09-078-294-4	Sequence 4, Appli
6	34.2	6.8	80595	US-09-078-294-3	Sequence 3, Appli
7	34	6.8	7218	US-08-233-463-14	Sequence 14, Appli
8	33	6.6	1859	US-08-691-563C-46	Sequence 46, Appli
9	32	6.4	1417	US-08-117-362-31	Sequence 31, Appli
10	32	6.4	1417	US-08-486-924-31	Sequence 31, Appli
11	32	6.4	1722	US-08-691-563C-58	Sequence 58, Appli
12	30.8	6.1	9763	US-08-973-273-1	Sequence 1, Appli
13	30.6	6.1	11827	US-09-733-455-3	Sequence 3, Appli
14	30.2	6.0	1043	US-09-422-576D-6	Sequence 6, Appli
15	30.2	6.0	1091	US-09-422-576D-5	Sequence 5, Appli
16	30.2	6.0	1165	US-09-422-576D-25	Sequence 25, Appli
17	30.2	6.0	1831	US-09-422-576D-1	Sequence 1, Appli
18	29.4	5.9	29598	US-09-341-587-6	Sequence 6, Appli
19	29.2	5.8	1289	US-08-896-164-16	Sequence 16, Appli
20	29.2	5.8	1132	US-08-894-731-3	Sequence 3, Appli
21	29	5.8	2023	US-08-961-083-199	Sequence 199, App
22	28.8	5.7	33768	US-08-961-083-71	Sequence 71, Appli
23	28.8	5.7	1066	US-08-793-559-2	Sequence 2, Appli
24	28.8	5.7	2142	US-08-793-559-3	Sequence 3, Appli
25	28.8	5.7	3568	US-09-218-363-3	Sequence 3, Appli
26	28.8	5.7	3825	US-08-737-597-1	Sequence 1, Appli
27	28.8	5.7	3825	US-08-737-597-2	Sequence 2, Appli

28	28.6	5.7	3552	4	US-09-157-210-3	Sequence 3, Appli
29	28.4	5.7	6836	4	US-08-976-259-73	Sequence 73, Appli
30	28.2	5.6	623	4	US-09-385-982-52	Sequence 52, Appli
31	28.2	5.6	16389	4	US-09-741-154-3	Sequence 3, Appli
32	28.2	5.6	19307	3	US-08-836-022A-10	Sequence 10, Appli
33	28.2	5.6	19307	4	US-09-427-048A-10	Sequence 10, Appli
34	28	5.6	2364	2	US-08-838-219B-5	Sequence 5, Appli
35	28	5.6	2364	3	US-09-233-336A-5	Sequence 5, Appli
36	28	5.6	2364	4	US-09-233-752A-5	Sequence 5, Appli
37	28	5.6	2364	4	US-09-402-036-5	Sequence 5, Appli
38	28	5.6	2364	4	US-09-904-326-5	Sequence 5, Appli
39	28	5.6	2370	4	US-09-002-285-79	Sequence 79, Appli
40	28	5.6	2370	4	US-09-002-285-95	Sequence 95, Appli
41	28	5.6	2374	4	US-09-002-285-97	Sequence 97, Appli
42	28	5.6	2375	4	US-08-960-780-3	Sequence 3, Appli
43	28	5.6	2375	4	US-08-960-780-7	Sequence 7, Appli
44	28	5.6	2375	4	US-09-073-898-3	Sequence 3, Appli
45	28	5.6	2375	4	US-09-073-898-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-09-175-928-3
Sequence 3, Application US/09175928A
Patent No. 6312921
GENERAL INFORMATION:
APPLICANT: Jacoby, Kenneth
APPLICANT: McCoy, John W.
APPLICANT: Lavallie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Metberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Mi, Sha
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: 6006B.AJ172A
CURRENT APPLICATION NUMBER: US/09/175, 928A
CURRENT FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-09-175-928-3
Query Match 74.9%; Score 375.4; DB 4; Length 2946;
Best Local Similarity 99.7%; Pred. No. 5.1e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 TCGGCAACCTCCCAACAGCAGCTTAGTTTCTCTGTGAGATGGGAGCTGAGAGACAG 63
DB 2553 TCGGCAACCTCCCAACAGCAGCTTAGTTTCTCTGTGAGATGGGAGCTGAGAGACAG 2612
QY 64 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCTTACCTGGAGAGTGAACAC 123
DB 2613 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCTTACCTGGAGAGTGAACAC 2612
QY 124 ATTCACCTTTAAACAGGGGCTTCACTAGCTACACCTGACCAATCAGAGAGCTCAC 183
DB 2673 ATTCACCTTTAAACAGGGGCTTCACTAGCTACACCTGACCAATCAGAGAGCTCAC 2732
QY 184 TAAATGCTAATTAGGCAAGAGAGAGATGAAGAAATAGCAATCATCTATTGCTGAG 243
DB 2733 TAAATGCTAATTAGGCAAGAGAGAGATGAAGAAATAGCAATCATCTATTGCTGAG 2792
QY 244 AGCAGAGAGAGAGAGAGATGATGAGATATAACCCCAAGTCTTGAGCCGGCAACGGCA 303
DB 2793 AGCAGAGAGAGAGAGAGATGATGAGATATAACCCCAAGTCTTGAGCCGGCAACGGCA 2852
```

QY 304 ACCCCCTTGGGTCCTCCCTCTTGTATGAGAGCT-TGTTTTCATGCTATTTCACCTAT 363
DB 2853 ACCCCCTTGGGTCCTCCCTCTTGTATGAGAGCT-TGTTTTCATGCTATTTCACCTAT 2912
QY 364 TAAATCTTGCACTGCA 380
DB 2913 TAAATCTTGCACTGCA 2929

RESULT 2

US-08-686-878A-50
Sequence 50, Application US/08686878A
Patent No. 5708157
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Evans, Cheryl
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/686,878A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-686-878A-50

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

QY 119 ACCACATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAG 178
DB 1 RCACATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAG 60
QY 179 CTCACATAATGCTAATTTAGGCAAGACGAGGTAAATAATAGCAATCATCTATTGC 238
DB 61 NTCANTAAATGATATATTGCGCAAAAACGAGGTAAATAATAGCAATCATCTATTGC 120
QY 239 CTGAGAGCAGACGAGGAGGACAATGATCGGATATTAACCAAGCTTCGAGCGGCAA 298
DB 121 CTGAGAGCAGACGAGGAGGACAATGATCGGATATTAACCAAGCTTCGAGCGGCAA 180
QY 299 CGGCAACCCCTTTGGGTCCCTCTTTGTATGGAG-TCTGTTTTCATGCTATTTCAC 358

DB 181 CGGCAACCCCTTTGGGTCCCTCTTTGTATGGAGCTTGTTCATGCTATTTCAN 240
QY 359 TCTATTAATCTTGCACTGCA 380
DB 241 TATTAATAATNTGCACTGCA 262

RESULT 3

US-08-721-489-4
Sequence 4, Application US/08721489
Patent No. 5786465
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John
APPLICANT: Lavallee, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/721,489
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-721-489-4

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

QY 119 ACCACATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAG 178
DB 1 RCACATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAG 60
QY 179 CTCACATAATGCTAATTTAGGCAAGACGAGGTAAATAATAGCAATCATCTATTGC 238
DB 61 NTCANTAAATGATATATTGCGCAAAAACGAGGTAAATAATAGCAATCATCTATTGC 120
QY 239 CTGAGAGCAGACGAGGAGGACAATGATCGGATATTAACCAAGCTTCGAGCGGCAA 298
DB 121 CTGAGAGCAGACGAGGAGGACAATGATCGGATATTAACCAAGCTTCGAGCGGCAA 180
QY 299 CGGCAACCCCTTTGGGTCCCTCTTTGTATGGAGCTTGTTCATGCTATTTCAN 358
DB 181 CGGCAACCCCTTTGGGTCCCTCTTTGTATGGAGCTTGTTCATGCTATTTCAN 240

APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,362
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 783..890
US-08-117-362-31

Query Match 6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGGCTGACTAAGATCCCTAAGCCTGAGGAGTGACCAATCCACC 130
Db 1345 TGCCCTTTTCAATTGGGCTAATCTCTTGAGGTCAATGCTAAGCACTAAAGAAAA 1286

QY 131 TTTAAACCGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGAGCTCACTAAATG 190
Db 1285 TTTAAGAACTGGGTTTAAAAAAATCTCATTTACAAATGAAGAGCTGCATT 1226

QY 191 CTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCATC---ATCTATTGCTGAGAGC 246
Db 1225 CAATGTGACCAACAACAGAGTAAATAGTGGGTATCTTTTCTCTTTTAAAAAC 1166

QY 247 ACAGCAGAGGAGCAATGATCGGATATATAACCCAGTCTTGAGACCGCAACGGCAAC 306
Db 1165 AGACCATGAATGAGAAATGAAGCAAGAACACTGCTGAGAACTGAAGTAAAGA 1106

QY 307 CCGTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTTCATCTATTA 366
Db 1105 AATATCTGTTTCTTAATGTTTGAAGGATTTGAATTTGTTTCATTTCACATCAGCA 1046

QY 367 ATCTTGCACTGCACTCTTGTGTCATGTTTCTTACG 404
Db 1045 TGCTTTCAACAACATTCCAAGTTACATGGTTTGAAG 1008

RESULT 10
US-08-486-924-31/C
; Sequence 31, Application US/08486924
; Patent No. 5789197
; GENERAL INFORMATION:
; APPLICANT: Wetterau II, John R.
; APPLICANT: Sharp, Daru Y.

APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 783..890
US-08-486-924-31

Query Match 6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGGCTGACTAAGATCCCTAAGCCTGAGGAGTGACCAATCCACC 130
Db 1345 TGCCCTTTTCAATTGGGCTAATCTCTTGAGGTCAATGCTAAGCACTAAAGAAAA 1286

QY 131 TTTAAACCGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGAGCTCACTAAATG 190
Db 1285 TTTAAGAACTGGGTTTAAAAAAATCTCATTTACAAATGAAGAGCTGCATT 1226

QY 191 CTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCATC---ATCTATTGCTGAGAGC 246
Db 1225 CAATGTGACCAACAACAGAGTAAATAGTGGGTATCTTTTCTCTTTTAAAAAC 1166

QY 247 ACAGCAGAGGAGCAATGATCGGATATATAACCCAGTCTTGAGACCGCAACGGCAAC 306
Db 1165 AGACCATGAATGAGAAATGAAGCAAGAACACTGCTGAGAACTGAAGTAAAGA 1106

QY 307 CCGTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTTCATCTATTA 366
Db 1105 AATATCTGTTTCTTAATGTTTGAAGGATTTGAATTTGTTTCATTTCACATCAGCA 1046

QY 367 ATCTTGCACTGCACTCTTGTGTCATGTTTCTTACG 404
Db 1045 TGCTTTCAACAACATTCCAAGTTACATGGTTTGAAG 1008

RESULT 11
US-08-691-563C-58
; Sequence 58, Application US/08691563C

Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: Herve PERRON
APPLICANT: Frederic BESEME
APPLICANT: Frederic BREDIN
APPLICANT: Glaucia PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-RADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTI-STEP SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1722 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-58
Query Match 6.4%; Score 32; DB 3; Length 1722;
Best Local Similarity 73.2%; Pred. No. 0.66;
Matches 41; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1 GTCTGGCCCAACCTCCCAACGACACTTGGTTTCCTGTTGAGATGGGGACTGA 56
Db 1024 GTGATCGGCCCAATTCCTCCCAACGAGAGTGGGTGCTCTGTTTGAAGGGGGGATGA 1079
RESULT 12
US-08-973-273-1/c
Sequence 1, Application US/08973273
Patent No. 6140085
GENERAL INFORMATION:
APPLICANT: Dean, Caroline
APPLICANT: MacKnight, Richard C
APPLICANT: Bancroft, Ian
APPLICANT: Lister, Clare K
TITLE OF INVENTION: Genetic Control of Flowering
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderyhe P.C.
STREET: 1100 No. 6140085th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,273
FILING DATE: 01-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/01332
FILING DATE: 03-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 951196.9
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-29
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9763 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Arabidopsis thaliana
STRAIN: Columbia
US-08-973-273-1
Query Match 6.1%; Score 30.8; DB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 4.4;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 162 CTTGACCAATCAGAGAGCTCTAAATGCTAATTAGGCAAGACAGAGGTAAGAAT 221
Db 4845 CCTTCAGAGGAAGAGAACTATATCATCTGATAGGCAAGAAACAGGAAAGAAA 4786
Qy 222 AGCCAATCATCTAT 235
Db 4785 GGATCAACATCCAT 4772
RESULT 13
US-09-739-455-3
Sequence 3, Application US/09739455
Patent No. 6413756
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL000653
CURRENT APPLICATION NUMBER: US/09/739,455
CURRENT FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 11827
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(11827)
OTHER INFORMATION: n = A,T,C or G
US-09-739-455-3
Query Match 6.1%; Score 30.6; DB 4; Length 11827;
Best Local Similarity 51.1%; Pred. No. 5.8;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy	103	CCTACTGGGAAGGAGACCAATCCACCCTTTAAACACGGGGCTTGCAACTTAGCTCACAC	162
Db	10853	CCCACTGTGTGAGAGTCGGAACGCTGCGCTTTTGTGGCGCGCTTGTTCTTAAATCAGTT	10912
Qy	163	CTGACCAATCAGAGAGCTCACTAAATGCTAATTAGGCANNAAGACAGAGGTAAAGAAATA	222
Db	10913	CCCTCTTTCAGATTTTATTAACACTAAAAAATAATAGTTTGTAAAAAATAATGAGAAATA	10972
Qy	223	GCCATCATCTATTGCCTGAG	243
Db	10973	CAGAAACATGAATTTACAGAG	10993

RESULT 14
US-09-422-576D-6

```

1  TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use
2  TITLE OF INVENTION: Recombinant Vectors
3  FILE REFERENCE: M0351-205010
4  CURRENT APPLICATION NUMBER: US/09/422,576D
5  CURRENT FILING DATE: 1998-10-21
6  PRIOR APPLICATION NUMBER: US 60/105,256
7  PRIOR FILING DATE: 1998-10-22
8  NUMBER OF SEQ ID NOS: 25
9  SOFTWARE: PatentIn version 3.1
10 SEQ ID NO 6
11
12 LENGTH: 1043
13
14 TYPE: DNA
15
16 ORGANISM: Homo sapiens
17
18 US-09-422-576D-6

```

	Query Match	Best Local Similarity	Score 30.2;	DB 4;	Length 1043;
	Matches 64; Conservative		Pred. No.2.1;	Matches 38; Indels 1;	Gaps 14
Qy	253	GGAGGGAACATGATCGGAGATATATAACCCAAAGCTTTGAGCGCGGCAACGGCAACCCCTTT	312		
Db	485	GGTGGGCGCCAGATTAAGAGATATAAGCAAGCTGCCCGGACCAAGCACTGGCAACCGGC-TC	543		
Qy	313	GGGTCCCTCCCTTTGTATGGAGAGCTGTGTTTCATGCAATTT	355		
Db	544	GGGTCCCTTCACACTGTGAAGCTTTGTCTTGGCTCTTT	586		

```

RESULT 15
US-09-422-576D-5
: Sequence 5, Application US/09422576D
: Patent No. 6395549
: GENERAL INFORMATION:
: APPLICANT: Tuan, Dorothy
: APPLICANT: Long, Qiaoming
: APPLICANT: Bengira, Chikh
: TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use i
: TITLE OF INVENTION: Recombinant Vectors
: FILE REFERENCE: M0351-205010
: CURRENT APPLICATION NUMBER: US/09/422,576D
: CURRENT FILING DATE: 1999-10-21
: PRIOR APPLICATION NUMBER: US 60/105,256
: PRIOR FILING DATE: 1998-10-22
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 5
: LENGTH: 1091
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus Sequence

```

```
; NAME/KEY: misc_feature
; LOCATION: (437)..(437)
; OTHER INFORMATION: n = any nucleotide
US-09-422-576D-5
```

	Query Match	Similarity	6.0%	Score 30.2	DB 4	Length 1091
	Match	Local Similarity	62.1%	Pred. No. 2.1		
	Matches	64	Conservative	0	Mismatches	38
					Indels	1
					Gaps	1
Qy	253	GGAGGACATGATGGGATATAAACCCAGCTTCGAGCGGCAACGCAACCCCTTT	312			
Db	534	GCTGGGGCCAGATTAGAGATATAAGCAGGCTGCCCGAGCCAGCAGTGGCAACCCGC	TC 592			
Qy	313	GGGTCCCTCCCTTGTATGGAGCTCTGTTTCATGCTATTT	355			
Db	553	GGGTCCCTTCCACACTGTGGAAGCTTTGTTCTTCCGCTCTTT	635			

Search completed: April 19, 2003, 14:49:59
Job time : 94.6897 secs

```

; NAME/KEY: misc feature
; LOCATION: (437)..(437)
; OTHER INFORMATION: n = any nucleotide
US-09-422-576D-5

Query Match          6.0%; Score 30.2; DB 4; Length 1091;
Best Local Similarity 62.1%; Pred. No. 2.1;
Matches 64; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 253 GGAGGGACATGATCGGGATATTAACCCAGTCTTCGAGCCGGCAACGGCAACCCCTTT 312
    |||||
DB 534 GGTGGGGCCAGATAAGAGATTAAGCAGGCTGCCCGAGCCAGCAGTGGCAACCCGC-TC 592
    |||||

QY 313 GGGTCCCTCCCTTTGTATGGAGCTGTGTTTCATGCTATTT 355
    |||||
DB 593 GGGTCCCTCCCTCCACACTGTGGAGGCTTGTCTTTGCTCTTT 635
    |||||

Search completed: April 19, 2003, 14:49:59
Job time : 94.6897 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:08:18 ; Search time 75.717 Seconds
(without alignments)
6646.925 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gtcgcggcaactcccca.....ctgatcatgaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	426	85.0	1329	7 US-08-979-847-108	Sequence 108, App
2	375.4	74.9	2946	9 US-10-114-893-134	Sequence 134, App
3	374.4	74.7	2930	10 US-09-802-535-1	Sequence 1, Appl
4	341.8	68.2	635	7 US-08-979-847-102	Sequence 102, App
5	317.6	63.4	180557	12 US-10-003-806-6	Sequence 6, Appl
6	317.6	63.4	180557	12 US-10-003-806-9	Sequence 9, Appl
7	311.4	62.2	410	10 US-09-880-107-538	Sequence 58, App
8	305.2	60.9	326014	10 US-09-731-231A-3	Sequence 3, Appl
9	286.5	57.2	541	10 US-09-864-761-8173	Sequence 8173, App
10	281.2	56.1	1894	10 US-09-864-761-4444	Sequence 4444, App
11	266.8	53.3	15425	9 US-10-091-504-1654	Sequence 1654, App
12	266.8	53.3	15425	9 US-09-764-869-1654	Sequence 1654, App
13	238.4	40.7	279	12 US-10-040-916-50	Sequence 50, Appl
14	203.8	40.7	569	10 US-09-864-761-14951	Sequence 14951, A
15	146.2	29.2	246	10 US-09-864-761-20462	Sequence 20462, A
16	136.8	27.3	446	10 US-09-811-284-3	Sequence 3, Appl
17	136.2	27.2	504	10 US-09-864-761-7027	Sequence 7027, App
18	129.6	25.9	3524	10 US-09-972-724-1	Sequence 1, Appl
19	124.8	24.9	475	10 US-09-864-761-895	Sequence 895, App

C 20	122.4	24.4	559	10 US-09-864-761-7501	Sequence 7501, App
C 21	111.6	22.3	440	10 US-09-864-761-3694	Sequence 3694, App
C 22	105.4	21.0	409	10 US-09-864-761-4153	Sequence 4153, App
C 23	92.2	18.4	579	10 US-09-864-761-13678	Sequence 13678, A
C 24	58.8	11.7	387	10 US-09-864-761-30194	Sequence 30194, A
C 25	58	11.6	525	10 US-09-893-737-31	Sequence 31, Appl
C 26	55.4	11.1	88	10 US-09-864-761-20907	Sequence 20907, A
C 27	50.6	10.1	275	10 US-09-864-761-17675	Sequence 17675, A
C 28	45.4	9.1	464	10 US-09-867-701-2947	Sequence 2947, App
C 29	42.6	8.5	929	10 US-10-093-766-14	Sequence 14, Appl
C 30	42.2	8.4	362	10 US-09-867-701-5411	Sequence 5411, App
C 31	41.8	8.3	1529	9 US-10-125-237-75	Sequence 75, Appl
C 32	41.2	8.2	81	10 US-09-864-761-24915	Sequence 24915, A
C 33	39.2	7.8	635	9 US-09-796-692-5087	Sequence 5087, App
C 34	39	7.8	3824	9 US-10-036-041-22	Sequence 22, Appl
C 35	39	7.8	3824	9 US-10-028-072-541	Sequence 541, App
C 36	39	7.8	3824	9 US-10-035-855-22	Sequence 22, Appl
C 37	39	7.8	3824	9 US-10-121-049-541	Sequence 541, App
C 38	39	7.8	3824	9 US-10-123-904-541	Sequence 541, App
C 39	39	7.8	3824	9 US-10-140-470-541	Sequence 541, App
C 40	39	7.8	3824	9 US-09-931-836-22	Sequence 22, Appl
C 41	39	7.8	3824	9 US-10-175-746-541	Sequence 541, App
C 42	39	7.8	3824	9 US-10-176-918-541	Sequence 541, App
C 43	39	7.8	3824	9 US-10-176-921-541	Sequence 541, App
C 44	39	7.8	3824	9 US-10-227-884-209	Sequence 209, App
C 45	39	7.8	3824	9 US-10-036-214-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-979-847-108
Sequence 108, Application US/08979847
Publication NO. US20030039664A1

GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMRIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUXE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979, 847
FILING DATE: 26-NOV-1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-856-6400
TELEFAX: 703-856-2787

INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:
LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-108

Query Match 85.0%; Score 426; DB 7; Length 1329;
Best Local Similarity 90.6%; Pred. No. 8.2e-136;
Matches 453; Conservative 0; Mismatch 47; Indels 0; Gaps 0;

QY 2 TCTGCGCAACCTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGCTGAGAGAC 61
DB TGTGCGCAACCTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGCTGAGAGAC 556
QY 62 AGGATTAAGTGAATTTCTAGAGCTGATGAATCTCTAGAGCTGAGAGAGTGAAC 121
DB AGGATTAAGTGAATTTCTAGAGCTGATGAATCTCTAGAGCTGAGAGAGTGAAC 616
QY 122 ACATCCACCTTTAAACAGGGGCTTGCAACTAGCTCAACCTGACCAATCAGAGAGCTC 181
DB GCATCACTTTTAAACAGGGGCTTGCAACTAGCTCAACCTGACCAATCAGAGAGCTC 676
QY 182 ACTAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 241
DB ACTAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 736
QY 242 AGAGCAAGAGAGAGAGAGAGATGATGAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 301
DB AGAGCAAGAGAGAGAGAGAGATGATGAGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 796
QY 302 CAACCCCTTTGGGATCCCTCCCTTTGATGGAGCTCTGTTTCAATGCTATTTCACTCT 361
DB CAACCCCTTTGGGATCCCTCCCTTTGATGGAGCTCTGTTTCAATGCTATTTCACTCT 856
QY 362 ATTAATCTTGCACTGCACTCTTCTGCTCATGTTCTTAAGGCTTGAAGCTGAGCTTTC 421
DB ATTAATCTTGCACTGCACTCTTCTGCTCATGTTCTTAAGGCTTGAAGCTGAGCTTTC 916
QY 422 GCTGCCATCCCACTGCTGTTTGGCCGACCGCAACCCGCCCTGACTCCCATCTCT 481
DB GCTGCCATCCCACTGCTGTTTGGCCGACCGCAACCCGCCCTGACTCCCATCTCT 976
QY 482 CTGATCATGAGGGTGTCC 501
DB CTGATCATGAGGGTGTCC 996

RESULT 2

US-10-114-893-134
Sequence 134, Application US/10114893
Publication No. US20020193567A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Bowman, Michael R.
APPLICANT: Carlin-Duckett, McKeough
APPLICANT: Kelleher, Kerry S.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYPEPTIDES ENCODING THEM
FILE REFERENCE: GI 6000-10A
CURRENT APPLICATION NUMBER: US/10/114,893
EARLIER FILING DATE: 2002-04-02
CURRENT FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 321

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 134
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 74.9%; Score 375.4; DB 9; Length 2946;
Best Local Similarity 99.7%; Pred. No. 3e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGCTGAGAGAC 63
DB TCGGCAACCTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGCTGAGAGAC 2612
QY 64 GACTAGCTGATTTCTAGAGCTGATGAATCTCTAGAGCTGAGAGAGTGAAC 123
DB GACTAGCTGATTTCTAGAGCTGATGAATCTCTAGAGCTGAGAGAGTGAAC 2672
QY 124 ATCCACCTTTAAACAGGGGCTTGCAACTAGCTCAACCTGACCAATCAGAGAGCTC 183
DB ATCCACCTTTAAACAGGGGCTTGCAACTAGCTCAACCTGACCAATCAGAGAGCTC 2732
QY 184 TAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 243
DB TAAATGCTAATTAGGCAAGAGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 2792
QY 244 AGCAGAGAGAGAGAGAGATGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 303
DB AGCAGAGAGAGAGAGAGATGAGGATTAAGGATTAAGGATTAAGGATTAAGGATTAAG 2852
QY 304 ACCCCCTTTGGGATCCCTCCCTTTGATGGAGCTCTGTTTCAATGCTATTTCACTCT 363
DB ACCCCCTTTGGGATCCCTCCCTTTGATGGAGCTCTGTTTCAATGCTATTTCACTCT 2912
QY 364 TAAATCTTGCACTGCA 380
DB TAAATCTTGCACTGCA 2929

RESULT 3

US-09-902-535-1
Sequence 1, Application US/09902535
Patent No. US20020102530A1
GENERAL INFORMATION:
APPLICANT: Keith, Jr., James C.
APPLICANT: McCoy, John M.
TITLE OF INVENTION: Methods and compositions for diagnosing
TITLE OF INVENTION: and creating preclampsia and gestational trophoblast
FILE REFERENCE: GIN-6006B4
CURRENT APPLICATION NUMBER: US/09/902,535
CURRENT FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: 60/216,657
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (930)...(2546)
US-09-902-535-1

Query Match 74.7%; Score 374.4; DB 10; Length 2930;
Best Local Similarity 99.7%; Pred. No. 6.6e-118;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAAGTTTCCTGTTGAGATGGGGAGCTGAGAGAC 63
|||||

Db 2555 TCGGCCAACTCCCAACAGCATTAGCTTTTCTGTTGAGTGGGGACTGAGAGACAG 2614
Qy 64 GACTAGCTGATTTTCTAGGCTGATGAATAATCCCTAAGCCTAGCTGGGAAGTACAC 123
Db 2615 GACTAGCTGATTTTCTAGGCTGATGAATAATCCCTAAGCCTAGCTGGGAAGTACAC 2674
Qy 124 ATCCACCTTTAAACACGGGGCTTGCACCTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
Db 2675 ATCCACCTTTAAACACGGGGCTTGCACCTTAGCTCAGACCTGACCAATCAGAGACTCAG 2734
Qy 184 TAAATGCTAATTAGGCAAAAGACAGAGATTAAGCAATAGCAATCATCTATGCTTAG 243
Db 2735 TAAATGCTAATTAGGCAAAAGACAGAGATTAAGCAATAGCAATCATCTATGCTTAG 2794
Qy 244 AGCAGACGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGCAGCCGCAACGGCA 303
Db 2795 AGCAGACGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGCAGCCGCAACGGCA 2854
Qy 304 ACCCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTCTTTCATGCTATTTCACTCTAT 363
Db 2855 ACCCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTCTTTCATGCTATTTCACTCTAT 2914
Qy 364 TAAATCTTGCAACTGC 379
Db 2915 TAAATCTTGCAACTGC 2930

RESULT 4

US-08-979-847-102
; Sequence 102, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMURIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TUXE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OLIVE & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 102:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-102

Query Match 68.2%; Score 341.8; DB 7; Length 635;
Best Local Similarity 94.2%; Pred. No. 4.6e-107;
Matches 355; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 4 TCGGCCAACTCCCAACAGCATTAGCTTTTCTGTTGAGTGGGGACTGAGAGACAG 63
Db 243 TCGGCCAACTCCCAACAGCATTAGCTTTTCTGTTGAGTGGGGACTGAGAGACAG 302
Qy 64 GACTAGCTGATTTTCTAGGCTGATGAATAATCCCTAAGCCTAGCTGGGAAGTACAC 123
Db 303 GACTAGCTGATTTTCTAGGCTGATGAATAATCCCTAAGCCTAGCTGGGAAGTACAC 362
Qy 124 ATCCACCTTTAAACACGGGGCTTGCACCTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
Db 363 ATCCACCTTTAAACACGGGGCTTGCACCTTAGCTCAGACCTGACCAATCAGAGACTCAG 422
Qy 184 TAAATGCTAATTAGGCAAAAGACAGAGATTAAGCAATAGCAATCATCTATGCTTAG 243
Db 423 TAAATGCTAATTAGGCAAAAGACAGAGATTAAGCAATAGCAATCATCTATGCTTAG 482
Qy 244 AGCAGACGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGCAGCCGCAACGGCA 303
Db 483 AGCAGACGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGCAGCCGCAACGGCA 542
Qy 304 ACCCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTCTTTCATGCTATTTCACTCTAT 363
Db 543 ACCCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTCTTTCATGCTATTTCACTCTAT 602
Qy 364 TAAATCTTGCAACTGCA 380
Db 603 TAAATCTTGCAACTGCA 619

RESULT 5

US-10-003-806-6/c
; Sequence 6, Application US/10003806
; Patent No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulrik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; PRIOR FILING DATE: 2001-11-02
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
; US-10-003-806-6

Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.9e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;

Qy 52 ACTGAGAGACGAGACTAGCTGATTTTCTAGGCTGATGAATAATCCCTAAGCCTAGCTGG 111
Db 58991 AGTGAAGACAGGACTAGCTGATTTTCTAGGCTGATGAATAATCCCTAAGCCTAGCTGG 58932
Qy 112 GAAGGTGACACATCACTTAATAACAGGGGCTTGCACCTTAGCTCAGACCTGACCAAT 171
Db 58931 GAAGGTGACCGCTTCCACCTTTAAACATGGGGCTTGCACCTTAGCTCAGACCGACCAAT 58972
Qy 172 C-----AGAGAGCTCAATAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATA 222

Dh 58871 CAGATAGTAAAGAGCTCACTAAATGCTAATTAGGCAAAAAAGAGAGTAAAGAAATA 58812
Qy 223 GCCATCATCTATTGCTGAGAGACAGAGAGGAGGACATGATCGGATATATAACCCAA 282
Dh 58811 GCCATCATCTATTGCTGAGAGACAGAGAGGAGGAGGACATGATCGGATATATAACCCAA 58752
Qy 283 GTCTTCGAGCCGCGCAACGCAACCCCTTTGGGTTCCTCCCTTTGTATGGAGGCTCTGT 342
Dh 58751 GCATCTGAGCCGCAACGAGCTACGCTTTGGGTTCCTCCCTTTGTATGGAGGCTCTGT 58692
Qy 343 TTTCAATGCTATTCTACTCTATTAATTTTGGACAGCA--CTCTCTGTGCTCCATGTTTCT 400
Dh 58691 -----CTTCACTATTAATCTTGAAGCTGCACTCTCTTTTGGTCTAATTTGT 58642
Qy 401 TACGCGTGAAGCTGAGCTTTCGCTGCCATCCACACTGCTGTTGGCCGACGAGAC 460
Dh 58641 CATGCTTGAAGCTGAGCTTTCCTGCGCGTCCACACTGCTGTTGGCGCTGTCCGACAC 58582
Qy 461 CCGCGGCTGACTCCCATCCCTCTGATCATGAGGAGTGTG 500
Dh 58581 CTGCTGCTGACTTCCATCCGTCGATCCGCAAGCTGTC 58542

RESULT 6
US-10-003-806-9/c

; Sequence 9, Application US/10003806
; Patent No. US2002011929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulmik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02060US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 9
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
; US-10-003-806-9

Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.5e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;

Qy 52 ACTGAGAGACAGAGCTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAAGCTAGCTGG 111
Dh 58991 AGTGAAGAGCAGAGCTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAAGCTAGCTGG 58932
Qy 112 GAAGGTGACCATCACTTTAAACAGGGGCTTGCACCTTAGCTCACACTGACCAAT 171
Dh 58931 GAAGGTGACCGCTTCCACTTTAAACATGGGGCTTGCACCTTAGCTCACACCCGACCAAT 58872
Qy 172 C-----AGAGAGCTCAATAATGCTAATTAGCAAAAGACAGAGGTAAAGAAATA 222
Dh 58871 CAGATAGTAAAGAGCTCACTAAATGCTAATTAGCAAAAGACAGAGGTAAAGAAATA 58812
Qy 223 GCCATCATCTATTGCTGAGAGACAGAGAGGAGGACATGATCGGATATATAACCCAA 282
Dh 58811 GCCATCATCTATTGCTGAGAGACAGAGAGGAGGACATGATCGGATATATAACCCAA 58752
Qy 283 GTCTTCGAGCCGCGCAACGCAACCCCTTTGGGTTCCTCCCTTTGTATGGAGGCTCTGT 342
Dh 58751 GCATCTGAGCCGCAACGAGCTACGCTTTGGGTTCCTCCCTTTGTATGGAGGCTCTGT 58692
Qy 343 TTTCAATGCTATTCTACTCTATTAATTTTGGACAGCA--CTCTCTGTGCTCCATGTTTCT 400
Dh 58691 -----CTTCACTATTAATCTTGAAGCTGCACTCTCTTTTGGTCTAATTTGT 58642

Qy 401 TACGCGTGAAGCTGAGCTTTCGCTGCCATCCACACTGCTGTTTGGCCGACGAGAC 460
Dh 58641 CATGCTTGAAGCTGAGCTTTCCTGCGCGTCCACACTGCTGTTTGGCGCTGTCCGACAC 58582
Qy 461 CCGCGGCTGACTCCCATCCCTCTGATCATGAGGAGTGTG 500
Dh 58581 CTGCTGCTGACTTCCATCCGTCGATCCGCAAGCTGTC 58542

RESULT 7
US-09-880-107-538/c

; Sequence 538, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 538
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
; US-09-880-107-538

Query Match 62.2%; Score 311.4; DB 10; Length 410;
Best Local Similarity 92.6%; Pred. No. 1e-96;
Matches 349; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

Qy 4 TCGGCAACCTCCCAACAGACCTTAGGTTTCTGTTGAGATGAGGGAGCTGAGAGACAG 63
Dh 377 TCAGCAACCTCCCAACAGACCTTAGGTTTCTGTTGAGAGAGGGAGCTGAGAGACAG 318
Qy 64 GACTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAAGCTAGCTGAGGAGTGACAC 123
Dh 317 GACTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAAGCTAGCTGAGGAGTGACAC 258
Qy 124 ATCCACTTTAAACAGGGGCTTGCACCTTAGCTCACACCTGACCAATCAGAGACTAC 183
Dh 257 ATCCACTTTAAACAGGGGCTTGCACCTTAGCTCACACCTGACCAATCAGAGACTAC 198
Qy 184 TAAATGCTAATTAGGCAAAAGAGAGGTAAAGAAATAGCCATTCATTTGCTGTA- 242
Dh 197 TAAATGCTAATTAGGCAAAAGAGAGGTAAAGAAATAGCCATTCATTTGCTGTA 138
Qy 243 GAGCAGACAGAGGAGCAATGATCGGATATTAACCAAGCTTTCGAGCCGAC-ACGG 301
Dh 137 GAGCAGATGGAGGAGCAAGAGATTGCAATTAACCAAGCATTTCCAGCCAGGAAAG 78
Qy 302 CAACCCCTTTGGGTCCCTCTCTTTGTATGGAGCTGTGTTTCATGCTATTTCACTCT 361
Dh 77 CAACCGCTTTGGGTCCCTCTCTTTGTATGGAGCTGTGTTTCATGCTATTTCACTCT 18
Qy 362 ATTAAATCTTGCAACTG 378
Dh 17 ATTAAATCTTGCAACTG 1

RESULT 8
US-09-731-231A-3/c
; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1

GENERAL INFORMATION: Karl et al
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: C001007
CURRENT APPLICATION NUMBER: US/09/731,231A
CURRENT FILING DATE: 2000-12-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 326014
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) - (326014)
OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match Best Local Similarity 60.9%; Score 305.2; DB 10; Length 326014;
Matches 344; Conservative 0; Mismatches 33; Indels 9; Gaps 1;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTTGAATGGGGAGTGAAGACAG 63
Db 170580 TCGGCACTCCCTCCCAACAGCACTTAGGTTTCTGTTGAATGGGGAGTGAAGACAG 170521
QY 64 GACTAGCTGATTTCTCTAGGCTGATAGAAATCCCTAAGCTAGTGGGAAGTACCAAC 123
Db 170520 GACTAGCTGATTTCTCTAGGCTGATAGAAATCCCTAAGCTAGTGGGAAGTACCTGC 170461
QY 124 ATCCACTTTAAACAGCGGGGCTTGCACTTAGCTACACCTGACCAATC-----AG 174
Db 170460 TTCTACCTTTAAACCGGGGCTTGCACTTAGCTACACCTGACCAATCAGGTAGGAAG 170401
QY 175 AGAGCTCAATAATGCTAATTAGCAAGAAGAGGTAAGAAATAGCAATCATCTA 234
Db 170400 AGAGCTCAATAATGCTAATTAGCAAGAAGAGGTAAGAAATAGCAATCATCTA 170341
QY 235 TTGCTGAGACAGCAGGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGCCG 294
Db 170340 TTGCTGAGACAGCAGGAGGAGCAATGATCGGATATTAACCCAGTCTTGAGCCG 170281
QY 295 GCAAGCGCAACCCCTTGGGTCCTCCCTTGTATGGGAGCTGTGTTTCATGCTATT 354
Db 170280 GCAATGCTACCATTTTGGTCCCTCTTGTATGGGAGCTGTGTTTCATGCTATT 170221
QY 355 TCACCTATTAAATCTTCAACTGCA 380
Db 170220 TCACCTATTAAATCTTCAACTGCA 170195

RESULT 9
US-09-864-761-8173/C
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wenheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 8173
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC01663.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
US-09-864-761-8173

Query Match Best Local Similarity 57.2%; Score 286.6; DB 10; Length 541;
Matches 383; Conservative 0; Mismatches 59; Indels 21; Gaps 4;

QY 49 GGAAGTGAAGACAGAGTACCTGATTTCTAGGCTGATTAAGAAATCCCTAAGCTTAC 108
Db 536 GGAAGTGAAGACAGAGTACCTGATTTCTAGGCTGATTAAGAAATCCCTAAGCTTAC 477
QY 109 TGGGAAGGTGACACATCCACTTAACACAGGGGCTTGAACCTAGTCAACCTGACC 168
Db 476 TGGGAAGGTGACACATCCACTTAACACAGGGGCTTGAACCTAGTCAACCTGACC 417
QY 169 AATC-----AGAGAGCTCAATAAATGCTAATTAAGCAAGAGCAAGAGTAAAGA 219
Db 416 AATCAGCTAGTAAGAGACAGCTCACTAAGGCTAATTCGCTAAGAGAGTAAATA 357
QY 220 ATAGCAATCATCTATTGCTGAGAGCAGCAGAGGAGCAATGATCGGATATTAAC 279
Db 356 ATAGCAATCATCTATTGCTGAGAGCAGCAGAGGAGCAATGATCGGATATTAAC 297
QY 280 CAAGCTTTGAGCGGAGCAAGGCAACCCCTTGGGTCCTCCCTTGTATGGAGGCTC 339
Db 296 CAGGATTTCCAGCGAGCAATGTAAGGCTTGAAGGTCCTCCACCTGATGAGGAGCTC 237
QY 340 TGTTCATGCTATTGCTACTTATTAATTTGCAACTGCACTCT-TCTGATCATGTTT 398

Db 236 TGT-----TTTCACTATTAATCTTG-2ACTGCACACTCTGCTGTGTTT 187
Qy 399 CTTAGCGTTGAGCTGACCTTTCGTCGCCATCAACCACTGTTTCCGCCACCGAG 458
Db 186 ATTCCGGTTCAATGAGCTTTTCTCAACATCAACACTGCTAGTACTGTCTGAG 127
Qy 459 AC-CGCGCTGACTCCATCCCTCTGATCATTCAGGGTTC 500
Db 126 ACACCGCTGCTGACTTCCACCCCTCGGATCTGCAGGGTTC 84
RESULT 10
US-09-864-761-4444
Sequence 4444, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4444
LENGTH: 1894
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC002346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444
Query Match 56.1%; Score 281.2; DB 10; Length 1894;
Best Local Similarity 88.4%; Pred. No. 5,4e-86;
Matches 320; Conservative 0; Mismatches 33; Indels 9; Gaps 1;
Qy 2 TCTGGCCCAACCTTCCCAACAGCACTTAGTTCCTGTTGAGATGGGGGACTGAGAGAC 61
Db 1457 TGTGGCCCAACCTTCCCAACAGCAGTGGGTTTCTGTTGAGAGGGGGGACTGAGAGAC 1516
Qy 62 AGGACTAGCTGAGATTTCTAGCTGAGCTAGAAATCCCTAAGCTAGCTGGGAGGTGACC 121
Db 1517 AGGATTAATGATGATTTCTTAGCACTTAAGATCCCTAAGCTAGCTGGGAGGTGACC 1576
Qy 122 ACATCCACCTTTAAACAGGGGCTTGCACCTAGCTCACTGACCAATC----- 172
Db 1577 GCTTCCACCTTTAAACAGGGGCTTGCACCTAGCTCACTGACCAATGAGTACTAA 1636
Qy 173 AGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATATC 232
Db 1637 AGAGAGCTCACTAAATGCTAATTAGGCAAGAGAGGTAAGAAATAGCCATATC 1696
Qy 233 TATTGCTGAGAGCAAGCAGAGAGGCAATGATCGGATTAACCAATCTTCGAGC 292
Db 1697 TATTGCTGAGAGCAAGCAGAGAGGCAATGATCGGATTAACCAATCTTCGAGC 1756
Qy 293 CGGCAGAGGCAACCCCTTGGTCCCTCCCTTGTATGGAGCTGTGTTGATGCTA 352
Db 1757 CAGTACAGCTACCTCTTGGTCCCTCCCTTGTATGGAGCTGTGTTGATGCTA 1816
Qy 353 TT 354
Db 1817 TT 1818
RESULT 11
US-10-091-504-1654/c
Sequence 1654, Application US/10091504
Publication No. US2003005908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091,504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1654
Query Match 53.3%; Score 266.8; DB 9; Length 15425;
Best Local Similarity 88.1%; Pred. No. 1,4e-80;
Matches 317; Conservative 0; Mismatches 32; Indels 11; Gaps 2;
Qy 4 TCGGCCAACCTTCCCAACAGCACTTAGTTCCTGTTGAGATGGGGGACTGAGAGAG 63
Db 2049 TTGGCCAACTTCCCAACAGCACTTAGTTCCTGTTGAGAGGGGGGACT--GAGAGAG 1992
Qy 64 GACTAGCTGAGATTTCTAGGCTGAGTAAATCCCTAAGCTAGCTGGGAGGTGACAC 123
Db 1991 AACTAGCTGAGATTTCTAGGCTGAGTAAATCCCTAAGCTAGCTGGGAGGTGACAC 1932

QY	124	ATCCACCTTTAAACACGGGGCTTGCAACTAGCTCACACTGACCAATC-----AG	174
Db	1931	ATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACCCGACCAATAGGATTTAAAG	1872
QY	175	AGAGCTCACTAAAATGCTAAATTAGGCAAAACAGAGAGTTAAAGAAATAGCCATCATCTA	234
Db	1871	AGAGCTCACAGAAATGCTAATTAGGCAAAAACAGAGAGTTAAACAAATAGCCATCATCTA	1812
QY	235	TTTGCTTAGAGCACACAGAGAGGGAACAATGATGGGATTTAAACCAAGCTTTGCAGCGC	294
Db	1811	TCGGCTTAGAGCACAGGGAAGGGAACAATGATGGGATTTAAACCAAGCATTTAGCTG	1752
QY	295	GCAACGGCAACCCCTTTGGGATCCCTCCCTTTGTATGAGAGCTGTGTTTCATGCTAAT	354
Db	1751	GCAACAGCTACCTCTTTGGATCCCTCCCTTTGTATAGAGCTGTGTTTTCATCTACT	1692

```

RESULT 12
US-09-764-869-1654/c
: Sequence 1654, Application US/09764869
: Patent No. US20020061521A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC007
: CURRENT APPLICATION NUMBER: US/09/764,869
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 2442
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1654
: LENGTH: 15425
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-869-1654

```

Query Match	53.3%;	Score 266.8;	DB 10;	Length 15425;
Best Local Similarity	88.1%;	Pred. No. 1.4e-80;		
Matches 317; Conservative	0;	Mismatches 32;	Indels 11;	Gaps 21

OY	4	TCGGCAACCTCCCCAAGACTAGTGTTCCTGTGAGATGGGGGACTAGAGACAG	63
Db	2049	TTGGCAACCTCCCCAAGACTAGTGTTCCTGTGAGAGGGGGGAGCT--GAGACAG	19929
OY	64	GACTAGCTGATTTCTTAGGCTGACTAGAAATCCCTAAGCCTAGCTGGGAGAGTACCAAC	123
Db	1991	AACTAGCTGATTTCTTAGTCCACTAAGAAATCCCTAAGCCTAGCTGGGAGAGTACTAC	19323
OY	124	ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCACCTGACCAATC-----AG	174
Db	1991	ATCCATCTTTAAACATGGGGCTTGCACTTAGCTCACCCGACCAATAGGATATTAAAG	18729
OY	175	AGAGCTCACTAAATGCTTAATTAGGCAAGACAGAGTAAAGAAATAGCCATATCTTA	234
Db	1871	AGAGCTCAAGAAATGCTTAATTAGGCAAAAACAGGAGTAAACAAATAGCCATATCTTA	18122
OY	235	TTGGCTGAGACACAGACAGAGGAGCAATATATGGGATATAAACCCAAATCTTCAGACG	294
Db	1811	TCGGCTTAGACACAGTGGGAGGAGCAATATGAGGATATAAACCCAGGACTTTAGCTG	17523
OY	295	GCAACGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGGAGCTGCTTTATCATCTAAT	354
Db	1751	GCAACAGTACCTCTTTGGGTCCCTCTTTGTATAGAGCTCTGTTTTACCTACT	16929

RESULT 13
US-10-040-916-50
; Sequence 50, Application US/10040916
; Patent No. US2002014679A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallie, Edward
;

1 Racie, Lisa
 2 Merberg, David
 3 Treacy, Maurice
 4 Evans, Cheryl
 5 Spaulding, Vikki
 6
 7 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 8 ENCODING THEM
 9
 10 NUMBER OF SEQUENCES: 71
 11
 12 CORRESPONDENCE ADDRESS:
 13 ADDRESSEE: Genetics Institute, Inc.
 14 STREET: 87 Cambridgepark Drive
 15 CITY: Cambridge
 16 STATE: Massachusetts
 17 COUNTRY: U.S.A.
 18
 19 ZIP: 02140
 20
 21 COMPUTER READABLE FORM:
 22 MEDIUM TYPE: floppy disk
 23 COMPUTER: IBM PC compatible
 24 OPERATING SYSTEM: PC-DOS/MS-DOS
 25 SOFTWARE: PatentIn Release #1.0, Version #1.30
 26
 27 CURRENT APPLICATION DATA:
 28 APPLICATION NUMBER: US/10/040,916
 29 FILING DATE: 07-Jan-2002
 30 CLASSIFICATION: <Unknown>
 31
 32 PRIOR APPLICATION DATA:
 33 APPLICATION NUMBER: 08/887,029
 34 FILING DATE: 07-FEB-1997
 35 APPLICATION NUMBER: 08/686,878
 36 FILING DATE: <Unknown>
 37
 38 ATTORNEY/AGENT INFORMATION:
 39 NAME: Brown, Scott A.
 40
 41 REGISTRATION NUMBER: 32,724
 42 TELECOMMUNICATION INFORMATION:
 43 TELEPHONE: (617) 498-8224
 44 TELEFAX: (617) 876-5851
 45
 46 INFORMATION FOR SEQ ID NO: 50:
 47 SEQUENCE CHARACTERISTICS:
 48 LENGTH: 279 base pairs
 49 TYPE: nucleic acid
 50 STRANDEDNESS: double
 51 TOPOLOGY: linear
 52
 53 MOLECULE TYPE: cDNA
 54
 55 SEQUENCE DESCRIPTION: SEQ ID NO: 50:
 56
 57 US-10-040-916-50

Query Match	47.6%	Score 238.4	DB 12	Length 279
Best Local Similarity	92.4%	Pred. No. 9.9e-72		
Matches 242	Conservative 1	Mismatches 19	Indels 0	Gaps 0
QY 119	ACCAATCCACCTTTAAACACGGGCGCTTGCACTTAGCTCAACCTGACCAATCAGAGAG	178		
Db 1	RCACATCCACCTTTAAACACGGGSGNTTGC AAANAAGATMACCTTGACCATCAGAGAG	60		
QY 179	CTCACTAAATGCTATTATTAGGCAAAAGACAGAGGTAAAGAAATAGCCCAATCATATTGGC	238		
Db 61	NTCANTAAATGATGATTATTGNGCAAAAAACAGAGGTAAAGAAATAGCCATCATATTATGGC	120		
QY 239	CTGAGAGCACAGCAGGAGGAGGACATGATCGGATATTAACCAAGCTTTTCAGCGCGCAA	298		
Db 121	CTGAGAGCACAGCAGGAGGAGGACAAATGATCGGATATTAAACCAAGTTTNGAGCGCGCAA	180		
QY 299	CGGCAACCCCTTTGGGTCGCCCTCCCTTTGATGGGAGCTCGTTTTCATGTATTTTAC	358		
Db 181	CGGCAACCCCTTTGGGTCGCCCTCCCTTTGATGGGAGCTGTTTTCATGTATTTTCAN	240		
QY 359	TCTATTAAATCTTGCACCTGCA	380		
Db 241	TNTATTAAATNTTGCACCTGCA	262		

RESULT 14
US-09-864-761-14951/C
; Sequence 14951, Application US/09864761

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14951
LENGTH: 569
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP00233.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER. SIGNAL = 9.3
OTHER INFORMATION: EXPRESSED IN PLACENTA. SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW. SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER. SIGNAL = 2.1
US-09-864-761-14951
Query Match 40.7%; Score 203.8; DB 10; Length 569;
Best Local Similarity 79.3%; Pred. No. 1.1e-59;
Matches 337; Conservative 0; Mismatches 62; Indels 26; Gaps 7;

Qy 171 TC-----AGAGACTCACTAAATGCTATTAGGC-AAAACAGAGAGTAAAGAA 220
Db 295 TCAGTAGTAAAGAGGGTTCCTAAATATACAAATTAGCTAAAGCAGAGAGTAAAGAA 236
Qy 221 TAG-CCATCATCTATTGCTGAGAGCAGAGAGAGGAGCAATGATCGGATATTAAC 279
Db 235 TAGCAATCATATATGCTGCTTAACAGCAGAGGAGGAGCAATGATTTGGATTTAAAC 176
Qy 280 CAA--GTCTTGAGCCGCAACGCAACCCCTTTGGTCCCTCCCTTTGTATGGAGC 337
Db 175 CAGAGATTAACACGGAGAGTGGCAACCCCTTTGGGTCCCTCCCATGTATGGAGC 116
Qy 338 TCTGTTTTCATGCTATTTCATCTATTAAATCTTGCACTG--CACTTTGTGTCATG 395
Db 115 TCTGT-----TTTCACTCTGTAAATCTTGCAACTGTATCACTCTGTGTCAGTG 66
Qy 396 TTTCTTACGCTTGAGCTGAGCTTTCGTCGCATCCACCACTGCTGTGTCGCCCAACG 455
Db 65 TTTGTTCCGCTCAAGCTGAACCTTTTGCTCACCTCTTACCACTGCTGTGTCGCCCAACG 6
Qy 456 CAGAC 460
Db 5 CAGAC 1

RESULT 15
US-09-864-761-20462/c
Sequence 20462, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30

```
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20462
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010951.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: AB026898.1, EVALUAE 4.00e-57
; OTHER INFORMATION: EST HUMAN HIT: AI492055.1, EVALUAE 3.00e-49
; OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALUAE 7.30e-01
US-09-864-761-20462

Query Match      29.2%; Score 146.2; DB 10; Length 246;
Best Local Similarity 88.5%; Pred. No. 4.4e-40;
Matches 185; Conservative 0; Mismatches 13; Indels 11; Gaps 2;

QY 232 CTATTGCTGAGAGCAGACGAGAGGAGACATGATCGGATATTAACCCAGTCTTCGAG 291
Db 246 CTGTTGCTGAGAGCAGACGAGAGGAGACATTAATCAGATATAAACCCAGCATTCGAG 187
QY 232 CCGGCAAGGCAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTGTTTCATGCT 351
Db 186 CTGGCAACGTAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTCTAT----- 136
QY 352 ATTCACTTATTAAATCTTGCAACTGCACTCTTCTGGTCAATGTTCTTAACGCTTGAG 411
Db 135 -CTTCACTTATTAAATCTTGCAACTGTAATCTTCTGGTCCGTGTTGTAC-GCTTGAG 78
QY 412 CTGAGCTTTGCTGCGCATCCACCACTGC 440
Db 77 CTGAGCTTTGCTGCGCATCCACCACTGC 49
```

Search completed: April 19, 2003, 14:58:25
Job time : 257.717 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 11:45:23 ; Search time 1268.51 Seconds
(without alignments)
6396.429 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gctcggcaccactcccca.....ctggatcatgcagggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376	75.0	440	12	BE732673
2	372.8	74.4	494	9	AA781423
3	372.2	74.3	436	9	AI128526
4	365.8	73.0	415	9	AI128496
5	360.4	71.9	388	14	H01325
6	355.4	70.9	727	17	AG029908

C	7	348.4	69.5	422	14	NS3177	NS3177 yv56h11.s1
	8	345.2	68.9	771	13	BI087886	BI087886 602852690
	9	343.8	68.6	683	17	AG102951	AG102951 Pan trogl
C	10	339.6	67.8	385	14	NS5091	NS5091 yv43e03.s1
	11	337.8	67.4	458	14	R76086	R76086 y171b03.s1
C	12	332.4	66.3	522	10	AW971553	AW971553 EST383642
	13	329	65.7	363	14	T69704	T69704 yd13a03.s1
C	14	326.6	65.2	342	9	AA860368	AA860368 aj59c05.s
C	15	326.4	65.1	424	14	R27412	R27412 yh46d11.s1
C	16	324.4	64.8	342	14	T47345	T47345 yb10h02.s1
C	17	317.8	63.6	438	14	R77278	R77278 y175d06.s1
C	18	317.6	63.4	443	9	AA837267	AA837267 od26b10.s
C	19	312.6	62.4	433	9	AI1379210	AI1379210 tcd01g11.x
C	20	312.6	62.2	490	9	AI598135	AI598135 tnl4a10.x
C	21	311.4	62.2	410	9	AA250958	AA250958 ze07d10.s
C	22	311	62.1	425	9	AI570707	AI570707 tm79g09.x
C	23	309.4	61.8	446	9	AI393478	AI393478 tga45g04.x
C	24	307.2	61.3	921	17	BH189565	BH189565 ENTQ48TR
C	25	305.4	61.0	463	14	R68685	R68685 y14g06.s1
C	26	302.6	60.4	404	14	R27389	R27389 yh46a09.s1
C	27	296.2	59.1	328	9	AA729556	AA729556 nx58c05.s
C	28	293.6	58.6	485	10	AM511366	AM511366 h445h03.x
C	29	293	58.5	431	9	AA552941	AA552941 nk61a10.s
C	30	290.8	58.0	471	9	AA709471	AA709471 zf91h06.s
C	31	289.6	57.8	653	17	AG033781	AG033781 Pan trogl
C	32	287.8	57.4	701	17	AG126669	AG126669 Pan trogl
C	33	285	56.9	611	17	AG381711	AG381711 RPII1-16
C	34	284	56.7	712	17	AO892947	AO892947 HS 3131.B
	35	281.2	56.1	722	17	AG049481	AG049481 Pan trogl
	36	279.2	55.7	503	14	BM723292	BM723292 UI-E-BUO-
	37	279.2	55.7	530	14	BM682297	BM682297 UI-E-BUO-
C	38	278.4	55.6	470	9	AI074704	AI074704 ox83d05.s
C	39	277.6	55.4	446	9	AI288235	AI288235 q18a08.x
C	40	276.2	55.1	960	17	AO900343	AO900343 HS 3179.B
	41	273.2	54.5	893	14	BQ437925	BQ437925 AGENCOURT
	42	272.8	54.5	609	17	AG066901	AG066901 Pan trogl
	43	272.4	54.4	326	14	D29167	D29167 HUMNR203 Hu
	44	272.4	54.4	447	12	BF919416	BF919416 QVO-NT015
	45	271.8	54.3	681	10	AV722664	AV722664 AV722664

ALIGNMENTS

RESULT 1
LOCUS BE732673 440 bp mRNA linear EST 15-SEP-2000
DEFINITION 601571305F1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3925728 5', mRNA sequence.
ACCESSION BE732673
VERSION BE732673.1 GI:10146665
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 440)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LICW52 row: j column: 01
High quality sequence stop: 440.
Location/Qualifiers
1..440
/organism="Homo sapiens"

/db xref="taxon:9606"
/clone="IMAGE:3925728"
/clone_id="N1H_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 124 a 117 c 101 g 58 t

ORIGIN

Query Match 75.0%; Score 376; DB 12; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.8e-111;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 63
Db 65 TCGGCAACCTCCCAACAGCACTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 124

QY 64 GACTAGCTGATTTCTAGCTGACTAGAAATCCCAGCTAGCTGGAGAGTGAACAC 123
Db 125 GACTAGCTGATTTCTAGCTGACTAGAAATCCCAGCTAGCTGGAGAGTGAACAC 184

QY 124 ATCCACCTTTAAACACGGGGCTTGGCACTTACTGTCACCTACCAATCAGAGACTCAC 244
Db 185 ATCCACCTTTAAACACGGGGCTTGGCACTTACTGTCACCTACCAATCAGAGACTCAC 244

QY 184 TAAATGCTAATTAGGCAAGAGAGGTAAGAAATATGCAATCATATATGCTGAG 243
Db 245 TAAATGCTAATTAGGCAAGAGAGGTAAGAAATATGCAATCATATATGCTGAG 304

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGCTTGGAGCCGCAACGCA 303
Db 305 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGCTTGGAGCCGCAACGCA 364

QY 304 ACCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTCAATGCTATTTCACTCTAT 363
Db 365 ACCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTCAATGCTATTTCACTCTAT 424

QY 364 TAAATCTTGCAACTGC 379
Db 425 TAAATCTTGCAACTGC 440

RESULT 2
AA781423/c 494 bp mRNA linear EST 31-DEC-1998
LOCUS a136c03.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'
DEFINITION similar to contains PTR7.tl PIR7 repetitive element; mRNA
sequence.
ACCESSION AA781423
VERSION AA781423.1 GI:2840754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
UNPUBLISHED (1997)
CONTACT Contact: Robert Strausberg, Ph.D.
Email: cgaap@bimail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D. Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.lnl.gov/bbrp/image/home
Insert Length: 1645 Std Error: 0.00
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 475.
Location/Qualifiers
1. 494
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="1391428"
/clone_id="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer 15'
TGTTACCAATCTGAGAGGGAGGGCCCAATTTTCTTTTCTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 111 c 133 g 141 t 1 others

ORIGIN

Query Match 74.4%; Score 372.8; DB 9; Length 494;
Best Local Similarity 99.2%; Pred. No. 2.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 63
Db 382 TCGGCAACCTCCCAACAGCACTAGGTTTCTTGTGAGATGGGGAGCTGAGAGACAG 323

QY 64 GACTAGCTGATTTCTAGCTGACTAGAAATCCCAGCTAGCTGGAGAGTGAACAC 123
Db 322 GACTAGCTGATTTCTAGCTGACTAGAAATCCCAGCTAGCTGGAGAGTGAACAC 263

QY 124 ATCCACCTTTAAACACGGGGCTTGGCACTTACTGTCACCTACCAATCAGAGACTCAC 183
Db 262 ATCCACCTTTAAACACGGGGCTTGGCACTTACTGTCACCTACCAATCAGAGACTCAC 203

QY 262 TAAATGCTAATTAGGCAAGAGGTAAGAAATATGCAATCATATATGCTGAG 243
Db 202 TAAATGCTAATTAGGCAAGAGGTAAGAAATATGCAATCATATATGCTGAG 143

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGCTTGGAGCCGCAACGCA 303
Db 142 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGCTTGGAGCCGCAACGCA 83

QY 304 ACCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTCAATGCTATTTCACTCTAT 363
Db 82 ACCCCTTTGGGTCCTTCTTGTATGGAGCTCTGTTTCAATGCTATTTCACTCTAT 23

QY 364 TAAATCTTGCAACTGC 380
Db 22 TAAATCTTGCAACTGC 6

RESULT 3
A1128526/c 436 bp mRNA linear EST 27-OCT-1998
LOCUS g661h10.x1 Soares_Placenta 8c9weeks 2bDH81c9w Homo sapiens cDNA
DEFINITION clone IMAGE:1714147 3' similar to contains PTR7.b1 PIR5 repetitive
element; mRNA sequence.
ACCESSION A1128526
VERSION A1128526.1 GI:3597040
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 436)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 720 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 428.
Location/Qualifiers
source
1. .436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1714147"
/clone_id="Soares_placenta_8to9weeks_2NDHP8to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Faelima Bernaldo."

BASE COUNT 95 a 96 c 106 g 139 t

Query Match 74.3%; Score 372.2; DB 9; Length 436;
Best Local Similarity 99.2%; Pred. No. 3.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGAGCTGAGAGAG 63
DB 399 TCGGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGAGCTGAGAGAG 340
QY 64 GACTAGCTGATTTCTTGAAGTGAATCCTTAAGCTTGAAGTGAAGTGAAG 123
DB 339 GACTAGCTGATTTCTTGAAGTGAATCCTTAAGCTTGAAGTGAAGTGAAG 280
QY 124 ATCCACCTTTAAACGCGGGCTTGAAGTGAATCCTTAAGCTTGAAGTGAAG 183
DB 279 ATCCACCTTTAAACGCGGGCTTGAAGTGAATCCTTAAGCTTGAAGTGAAG 220
QY 184 TAAATGCTAATTAGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
DB 219 TAAATGCTAATTAGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 160
QY 244 AGCAGCAGAGGAGGACATATGCGGATATTAACCAAGTCTTGAGCCGCAACGCA 303
DB 159 AGCAGCAGAGGAGGACATATGCGGATATTAACCAAGTCTTGAGCCGCAACGCA 100
QY 304 ACCCCCTTTGGGTCCCTTCCTTTGTATGAGAGCTCTGTTTCAAGCTATTCACTTAT 363
DB 99 ACCCCCTTTGGGTCCCTTCCTTTGTATGAGAGCTCTGTTTCAAGCTATTCACTTAT 40
QY 364 TAAATCTGCAACTGCA 380
DB 39 TAAATCTGCAACTGCA 23

RESULT 4
A1128496 415 bp mRNA linear EST 27-OCT-1998
LOCUS A1128496/c
DEFINITION q61e08.x1 Soares_placenta_8to9weeks_2NDHP8to9W Homo sapiens cDNA
clone IMAGE:171418 3' similar to contains PTKr.Tl PTKr repetitive
element; , mRNA sequence.
ACCESSION A1128496

VERSION A1128496.1 GI:3597010
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 1184 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 413.
Location/Qualifiers
source
1. .415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1714118"
/clone_id="Soares_placenta_8to9weeks_2NDHP8to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M.Faelima Bernaldo."

BASE COUNT 94 a 96 c 107 g 116 t 2 others

Query Match 73.0%; Score 365.8; DB 9; Length 415;
Best Local Similarity 97.6%; Pred. No. 3.7e-108;
Matches 370; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GTCTCGGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGAGCTGAGAG 60
DB 379 GTCTCGGCAACTCCCAAGACGACTAGGTTTCTGTGAGATGGGAGCTGAGAG 320
QY 61 CAGACTAGCTGATTTCTTGAAGTGAATCCTTAAGCTTGAAGTGAAGTGAAG 120
DB 319 CAGACTAGCTGATTTCTTGAAGTGAATCCTTAAGCTTGAAGTGAAGTGAAG 260
QY 121 CACATCCACCTTTAAACGCGGGCTTGAAGTGAATCCTTAAGCTTGAAGTGAAG 180
DB 259 CACATCCACCTTTAAACGCGGGCTTGAAGTGAATCCTTAAGCTTGAAGTGAAG 200
QY 181 CACTAATGCTAATTAGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 240
DB 139 CACTAATGCTAATTAGCAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 140
QY 241 GAGAGCAGAGGAGGACATATGCGGATATTAACCAAGTCTTGAGCCGCAACGCA 300
DB 139 GAGAGCAGAGGAGGACATATGCGGATATTAACCAAGTCTTGAGCCGCAACGCA 80
QY 301 GCAACCCCTTTGGGTCCCTTCCTTTGTATGAGAGCTCTGTTTCAAGCTATTCACTC 360
DB 79 GCAACCCCTTTGGGTCCCTTCCTTTGTATGAGAGCTCTGTTTCAAGCTATTCACTC 20
QY 361 TATTAAATCTGCAACTGC 379
DB 19 TATTAAATCTGCAACTGC 1

RESULT 5

H01325/c 188 bp mRNA linear EST 19-JUN-1995
 LOCUS Y199601.61 Soares Placenta NB2HP: Homo sapiens cDNA clone
 DEFINITION IMAGE:147384 3', mRNA sequence.
 ACCESSION H01325
 VERSION H01325.1
 KEYWORDS GI:864258
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 388)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
 'M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, P., Trevaekis, E., Waterston,
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 TITLE JOURNAL
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 790
 High quality sequence scope: 346
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL, contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 790 Std Error: 0.00
 Seq primer: Promega -21ml3
 High quality sequence scope: 346.
 Location/Qualifiers
 1..388
 /organism="Homo sapiens"
 /db_xref="GDB:559031"
 /db_xref="taxon:9606"
 /clone="IMAGE:147384"
 /clone_1lb="Soares placenta NB2HP"
 /sex="Female"
 /dev_stage="placenta obtained at birth (full term)"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Placenta; Vector: pTZ19 (pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 AACTGGAAGATTGGCGCGCCGACGATTTTTTTTTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pTZ19 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo. "
 BASE COUNT 90 a 90 c 96 g 110 t 2 others
 ORIGIN
 Query Match 71.9%; Score 360.4; DB 14; Length 388;
 Best Local Similarity 97.8%; Pred. No. 2.1e-106;
 Matches 364; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 Oy 4 TCGGCCAACCCTCCCAACGACGACTTAGGTTTCCTGTTGAGATGGGGGACGTAGAGACAG 63
 Db 377 TCGGNAACCTCCCAANAGCACTTAGGTTTCCTGTTGAGATGGGGGACGTAGAGACAG 318
 Oy 64 GACTAGCTGAGATTCTCTAGGCTGACCTAAGATCCCTAGCTGAGGAGGTGACAC 123
 Db 317 GACTAGCTGAGATTCTCTAGGCTGACCTAAGATCCCTAGCTGAGGAGGTGACAC 258
 Oy 124 ATCCACCTTTAAACACGCGGCTTGCACTTAGCTCAGCCTGACCAATCAGAGAGCTCAC 183
 Db 257 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCAGCCTGACCAATCAGAGAGCTCAC 198
 Oy 184 TAAATGCTAATTAGGCAAAACAGAGAGGTAAAGAAATGCCAATCATCTATTGCTGAG 243
 Db 197 TAAATGCTAATTAGGCAAAACAGAGAGGTAAAGAAATGCCAATCATCTATTGCTGAG 138

Oy 244 AGCAGCAGGAGGAGACAAATGATCGGATATTAACCCAGTCTTGCAGCCGCAACGGCA 303
 Db 137 AGCAGCAGGAGGAGACAAATGATCGGATATTAACCCAGTCTTGCAGCCGCAACGGCA 78
 Oy 304 ACCGCTTTGGGCTCCCTCCCTCTTTGATGAGAGCTCTGTTTCAATGCTATTCACTCAT 363
 Db 77 AGCCCTTTGGGCTCCCTCCCTCTTTGATGAGAGCTCTGTTTCAATGCTATTCACTCAT 18
 Oy 364 TAAATCTGCAA 375
 Db 17 TAAATCTGCAA 6
 RESULT 6 727 bp DNA linear GSS 01-NOV-2001
 LOCUS AG029908
 DEFINITION Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.
 ACCESSION AG029908
 VERSION AG029908.1
 KEYWORDS GI:16556781
 SOURCE GSS.
 Pan troglodytes male lymphoblast DNA, clone_1lb: PTB Chimpanzee Male
 BAC library clone: PTB-002C04.R.
 ORGANISM
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 BAC end sequences of library PTB
 Unpublished
 2 (bases 1 to 727)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimpanzee@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.
 PRIMERS
 Sequencing: M13Rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1..727
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-002C04.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_1lb="PTB Chimpanzee Male BAC Library"
 BASE COUNT 170 a 215 c 175 g 166 t 1 others
 ORIGIN
 Query Match 70.9%; Score 355.4; DB 17; Length 727;
 Best Local Similarity 91.1%; Pred. No. 1.2e-104;
 Matches 411; Conservative 0; Mismatches 36; Indels 4; Gaps 3;
 Oy 52 ACTGAGACGAGACTAGCTGATTTCTTAGGCTGACCTAAGATCCCTAAGACCTAGCTGG 111
 Db 275 ATTGAGACGAGACTAGCTGATTTCTTAGGCTGACCTAAGATCCCTAAGACCTAGCTGG 334
 Oy 112 GAAGGTGACCAATCAGCTTTAAACAGCGGCTTGCACTTAGCTCAGACCTGACCAAT 171
 Db 335 GAAGGTGACCAATCAGCTTTAAACAGCGGCTTGCACTTAGCTCAGACCTGACCAAT 394
 Oy 172 CAGAGGCTCACTAATGCTAATTAGCAAAACAGAGAGGTAAAGAAATGCCAATCAT 231

	Db	395	CGAGAGCTCAGTAAATGCTAATTATGGCAAAACAGAGGTAAAGAATAGCCAAATAT	454
QY	232	CTATTGCCTGAGAGCACAGACAGAGAGGA	CAATGATTCGGGATATAAACCCAAGTCTTTGAG	291
Db	455	CTATTGCCTGAGAGCACAGTGGGAAGATTAAGATTCGGGA	TATAAACCAAGCATTTCAAG	514
QY	292	CCGGCAAGGACCAACCCCCCTTTGGGTTGCCCTCCCTTTGTAATGGAGACTCTGTTTTATGCT	351	
Db	515	CGAGCAAGGGCAACCCCCCTTTGGGT-CCTCCCTTGTAATGGAGACTCTGTTTTACACT	573	
QY	352	ATTTCACCTATAATCTTGAACA	CTGACCTTCCTGGGCAAGTGTCTTCAAGGCTTAG	411
Db	574	ATTTCACCTATAATCTTGAACA	CTGACCTTCCTGGGCAAGTGTGTGTTGTTGTAAGGCTTAG	633
QY	412	CTGAGCTTTGCTGCTGCATCCACCACTGCTGT-TT-GCCGCGCACCGACAGACC	CGCGCTGTA	470
Db	634	CTGAGCTTTGCTGCTGCATCCACCACTGCTGTGTTTGGCCGCGCTGCACAGCCGCGCTTGA	693	
QY	471	CTCCCATCTCTCTGATCATGCAAGGCTGCC	501	
Db	694	CTTCATCTCTCT--GATCCAGCAGGCTGTCC	722	
RESULT 7				
LOCUS	NS3177/c			
DEFINITION	NS3177	422 bp	mRNA	linear EST 28-JAN-1997
	iy56hl.1.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone			
	JM58.246789.3'			
	similar to contains PTR.1.t2 PIR7 repetitive element			
	; mRNA sequence.			
ACCESSION	NS3177			
VERSION	NS3177.1	GI:1194343		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 422)			
AUTHORS	Hallier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,			
	Chissee,S., Dietrich,N., Dubugue,T., Favellio,A., Gish,W., Hawkins			
	,M., Hulman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore			
	,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rolfing,T.,			
	Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisakis,E.,			
	Underwood,K., Weidmann,P., Waterston,R., Wilson,R. and Marra,M.			
	Generation and analysis of 280,000 human expressed sequence tags			
	Genome Res. 6 (9), 807-828 (1996)			
	97044478			
COMMENT	Contact: Wilson RK			
	Washington University School of Medicine			
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: est@watson.wustl.edu			
	This clone is available royalty-free through LLNL, contact the			
	IMAGE Consortium (infoimage.llnl.gov) for further information.			
	Insert Length: 938 Std Error: 0.00			
	Seq primer: ml3 -40 forward			
	High quality sequence stop: 301.			
	Location/Qualifiers			
	1..422			
FEATURES	/organism="Homo sapiens"			
source	/db_xref="Gene:3796035"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:246789"			
	/clone_lib="Soares fetal liver spleen INFLS"			
	/sex="male"			
	/dev_stage="20 week-post conception fetus"			
	/lab_host="DH10B (ampicillin resistant)"			
	/note="Organ: Liver and Spleen; Vector: pTR73D (pharmacia)			
	with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;			
	1st strand cDNA was primed with a Pac I - oligo(dT) primer			
	[5' AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3']			
	double-stranded cDNA was ligated to Eco RI adaptors			

(Pharmacía), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTR73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."									
BASE COUNT	95 a	95 c	111 g	120 t	1 others				
ORIGIN									
Query Match	69.5%	Score 348.4;	DB 14;	Length 422;					
Best Local Similarity	98.7%	Pred. No. 1.1e-102;							
Matches 372;	Conservative	0;	Mismatches 2;	Indels 3;	Gaps 2;				
OY	4	TOGGCAACCTCCCAACAGCACTTAGGTTTCCGTGTGAGATGGGGGAGTGAAGACAG	63						
Db	375	TOGGCAACCTCCCAACAGCACTTAGGTTTCCGTGTGAGATGGGGGAGTGAAGACAG	316						
OY	64	GACTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTAGCTGGGAAGTACCAC	123						
Db	315	GACTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTAGCTGGGAAGTACCAC	256						
OY	124	ATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACCTGACCAATCAGAGACTCAC	183						
Db	255	ATCCACCTTTAAACACGGGGCTTGCAACTTAGCTCACCTGACCAATCAGAGACTCAC	196						
OY	184	TAAATGCTATTGAGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG	243						
Db	195	TAAATGCTATTGAGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG	136						
OY	244	AGCACAGCAGGAGGACATGATCGGATATTAACCCAAAGCTTTCGACCCGCAACGGCA	303						
Db	135	AGCACAGCAGGAGGACATGATCGGATATTAACCCAAAGCTTTCGACCCGCAACGGCA	77						
OY	304	ACCCCTTTGGGTCCTCCCTTTGTATGAGGAGCTCTGTTTCACTAGCTATTACTCTAT	363						
Db	76	ACCCC--TTTGATCCCTCCCTTTGTATGAGGAGCTCTGTTTCACTAGCTATTACTCTAT	19						
OY	364	TAAATCTTGCAACTGCA	380						
Db	18	TAAATCTTGCAACTGCA	2						
RESULT 8									
LOCUS	BI087886	771 bp	mRNA	linear	EST 20-JUN-2001				
DEFINITION	602852690P1 NIH_MGC_10 Homo sapiens CDNA clone IMAGE:4993894 5',								
ACCESSION	BI087886								
VERSION	BI087886.1	GI:14506216							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.								
TITLE	1 (bases 1 to 771)								
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/								
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)								
	Unpublished (1999)								
	Contact: Robert Strausberg, Ph.D.								
	Email: cgabs-r@mail.nih.gov								
	Tissue Procurement: ATCC								
	CDNA Library Preparation: Life Technologies, Inc.								
	CDNA Library Arrayed by: Incyte Genomics, Inc.								
	DNA sequencing by: Incyte Genomics, Inc.								
	Clone distribution: MGC clone distribution information can be								
	found through the I.M.A.G.E. Consortium/LINL at:								
	http://image.llnl.gov								
	Plate: LAM11015	row: d	column: 23						
FEATURES	High quality sequence. stop: 762.								
source	Location/Qualifiers								
	1..771								
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/clone_image="IMAGE:4993894"								
	/clone_id="NIH_MGC_10"								

```
/cell_line="MGC36"  
/lab_host="DH108"  
/note="Organ: cervix; Vector: PCMV-Spore6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: oligo dt.  
Average insert size 1.5 kb. Library prepared by life  
Technologies."  
BASE COUNT      225 a      206 c      166 g      174 t  
ORIGIN  
Query Match      68.9%; Score 345.2; DB 13; Length 771;  
Best Local Similarity 97.4%; Pred. No. 2.6e-101;  
Matches 372; Conservative 0; Mismatches 8; Indels 2; Gaps 2;  
QY 1 GTCCTGGCCCACTCTCCCAAGACGACCTAGGTTTCTGT-TGAGATGGGGAGCTGAGAG 59  
DB 377 GTCTGCGGCAACCTCCCAAGACGACCTAGGTTTCTGTGAGATGGGGAGCTGAGAG 436  
QY 60 ACAGGACTAGCTGATTTCTAGGCTGACTAGAAATCCCTAAGCTTACTGGAAGTGA 119  
DB 437 ACAGGACTAGCTGATTTCTAGGCTGACTAGAAATCCCTAAGCTTACTGGAAGTGA 496  
QY 120 CCACATCCACTTTTAAACAGGGGCTTGCACTTACCTCACACCTGACCAATCAGAGAGC 179  
DB 497 CCACATCCACTTTTAAACAGGGGCTTGCACTTACCTCACACCTGACCAATCAGAGAGC 556  
QY 180 TCACATAAATGCTAATTTAGGCAAGACAGAGGTA-AAAGAAATAGCCATCATCTATTGC 238  
DB 557 TCACATAAATGCTAATTTAGGCAAGACAGAGGTA-AAAGAAATAGCCATCATCTATTGC 616  
QY 239 CTGAGAGCAGACAGAGAGGACATGATCGGATTAACCCAGCTTTCGAGCCGCA 298  
DB 617 CTGAGAGCAGACAGAGAGGACATGATCGGATTAACCCAGCTTTCGAGCCGCA 676  
QY 299 CGGCAACCCCTTTGGGTCCTCCCTTTGATGGAGCTGCTGTTTATGATTTTAC 358  
DB 677 CGGCAACCCCTTTGGGTCCTCCCTTTGATGGAGCTGCTGTTTATGATTTTAC 736  
QY 359 TCTATTAAATCTTGCACTGCA 380  
DB 737 TCTATTAAATCTTGCACTGCA 758  
RESULT 9  
AG102951      683 bp      DNA      linear      GSS 03-NOV-2001  
LOCUS      Pan troglodytes DNA, clone: PNB-106G16.F, genomic survey sequence.  
ACCESSION      AG102951  
VERSION      AG102951.1 GI:16723468  
KEYWORDS      GSS.  
SOURCE      Pan troglodytes male lymphoblast DNA, clone lib:PTB Chimpanzee Male  
BAC library clone:PTB-106G16.F.  
ORGANISM      Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.  
REFERENCE      1  
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Tokoki,Y., Matanabe,H. and Sakaki,Y.  
TITLE      BAC end sequences of library PTB  
JOURNAL      Unpublished  
AUTHORS      Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Tokoki,Y., Matanabe,H. and Sakaki,Y.  
DIRECT SUBMISSION  
Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimbee@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the Rad process and may have higher chance of  
clone tracking errors.  
PRIMERS  
Sequencing: -21M13
```

```
LIBRARY  
Vector : pKS145  
R.site 1 : SacI  
R.site 2 : SacI  
FEATURES  
source  
1..683  
/organism="Pan troglodytes"  
/db_xref="taxon:9598"  
/clone="PTB-106G16.F"  
/sex="male"  
/cell_type="lymphoblast"  
/clone_lib="PTB Chimpanzee Male BAC library"  
BASE COUNT      162 a      186 c      167 g      167 t      1 others  
ORIGIN  
Query Match      68.6%; Score 343.8; DB 17; Length 683;  
Best Local Similarity 89.8%; Pred. No. 7.1e-101;  
Matches 403; Conservative 0; Mismatches 42; Indels 4; Gaps 3;  
QY 52 ACTGAGACAGACGCTAGCTGATTTCTAGGCTGACTAAGATCCTAAGCTTACTG 111  
DB 108 ATTGAGAGAAAGACTAGCTGATTTCTAGGCTGACTAAGATCCTAAGCTTACTG 167  
QY 112 GAAGTACCAATCCACCTTTAAACAGGGGCTTGCACTTACTGACACCTGACCAAT 171  
DB 168 GAAGTACCTGATCCACTTTTAAACATGGGCTTGCACTGACTGACCTGACCAAT 227  
QY 172 CAGAGAGCTCAATAATGCTAATTTAGGCAAGACAGAGGTTAAAGAAATAGCCATCAT 231  
DB 228 CAGAGAGCTCAATAATGCTAATTTAGGCAAGACAGAGGTTAAAGAAATAGCCATCAT 286  
QY 232 CTATTGCTGAGACACAGCAGAGGGAATGATCGGATTAATACCCAGTCTTGAG 291  
DB 287 CTATTGCTGAGACACAGCAGAGGGAATGATCGGATTAATACCCAGTCTTGAG 346  
QY 292 CCGGCAAGGAAACCCCTTTGGTCCCTCCCTTTGATGGGAGCTGTTTCAAGCT 351  
DB 347 CCGGCAAGGAAACCCCTTTGGTCCCTCCCTTTGATGGGAGCTGTTTCAAGCT 404  
QY 352 ATTTCACCTAATAATCTTGCACTGCTCTGCTCCATGTTTCTTAAGCGCTGAG 411  
DB 405 ATTTCACCTAATAATCTTGCACTGCTCTGCTCCATGTTTCTTAAGCGCTGAG 464  
QY 412 CTGAGCTTTGCTCGCATCCACCACTGCTG-TTTCGCGCCACCGCAGACCCGCTGA 470  
DB 465 CTGAGCTTTGCTCGCATCCACCACTGCTGTTTGTTCATTTGACAGACCCAGCTGA 524  
QY 471 CTCGCATCCCTCTGATCATCAGAGGTG 499  
DB 525 CTCGCATCCCTCTGATCATCAGAGGTG 553  
RESULT 10  
NS5091/c      385 bp      mRNA      linear      EST 28-JAN-1997  
LOCUS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
DEFINITION      yv43e03.61 Soares fetal liver spleen INFLS Homo sapiens cDNA clone  
IMAGE:245500 3', mRNA sequence.  
ACCESSION      NS5091  
VERSION      NS5091.1 GI:1197970  
KEYWORDS      EST.  
SOURCE      human.  
ORGANISM      Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE      1 (bases 1 to 385)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiappelli,B.,  
Chisese,S., Dietrich,N., Dubuque,T., Favellio,A., Gish,M., Hawkins  
M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore  
B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,  
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevasakis,E.,  
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL      Genome Res. 6 (9), 807-828 (1996)
```


QY	304	ACCCCTTTGGGATCCCTCCCTTGTATGAGGAGCCTGTTTTCATGCTATTCACTTAT	363
Db	59	A-CCCCCTTTGGGATCCCTCCCTTGTATGAGGAGCCTGTTTTCATGCTATTCACTTAT	1
RESULT 12			
LOCUS	AW971553/c	522 bp	mRNA linear EST 01-JUN-2000
DEFINITION	EST183842	MAGE resequences, MAGL Homo sapiens cDNA,	mRNA sequence.
ACCESSION	AW971553		
VERSION	AW971553.1	GI:8161399	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
	1 (bases 1 to 522)		
	Hedge, P., Qi, R., Abernathy, K., Dhurap, S., Gaspard, R., Gay, C., Holt		
	, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and		
	Quackenbush, J.		
TITLE	Assessment of gene expression patterns in a model of colon tumor		
JOURNAL	metastasis using a 19,200 element cDNA microarray		
COMMENT	Unpublished (2000)		
	Contact: John Quackenbush		
	The Institute for Genomic Research,		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel: 301 838 3528		
	Fax: 301 838 0208		
	Email: johnq@igr.org		
	Plate: 292		
FEATURES	Seq primer: Forward.		
source	Location/Qualifiers		
	1..522		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_1ib="MAGE resequences, MAGL"		
	/note="Vector: pBluescript-SK"		
BASE COUNT	125 a 115 c 134 g 148 t		
ORIGIN			
Query Match	66.3%; Score 332.4; DB 10; Length 522;		
Best Local Similarity	93.0%; Pred. No. 3.1e-97;		
Matches	348; Conservative 0; Mismatches 26; Indels 0; Gaps 0.		
QY	4	TCGGCAACCTCCCAACAGCACTTATGATTTCTGTGTGATGAGGAGCTGAGACAG	63
Db	374	TGAGCAACTCACAACAGCACTTGGTTTCTGTGTGATGAGGAGCTGAGACAG	315
QY	64	GACTAGCTGATTTCTTAGGCTGACTAAGANTCCCTAAGCTTAGTGGAGAGTGACAC	123
Db	314	GACTAGCTGATTTCTTAGGCTGACTAAGANTCCCTAAGCTTAGTGGAGAGTGACAC	255
QY	124	ATCCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACTGACCAATGAGAGCTCAC	183
Db	254	ATCCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACTGACCAATGAGAGCTCAC	195
QY	184	TAAATGCTAATTAGGCACAAAGACAGAGGTAAAGAAATAGCCAAATCATATTGCTGAG	243
Db	194	TAAATGCTAATTAGGCACAAAGACAGAGGTAAAGAAATAGCCAAATCATATTGCTGAG	135
QY	244	AGCACAGCAGAGAGGACAAATGATCGGATATTAACCAAGTCTTTCAGCCGACAGGCA	303
Db	134	AGCACAGTGGAGAGGACAAAGATTTGCAATTAACCCAGCATTTGAGCAGCAGGCA	75
QY	304	ACCCCTTTGGGATCCCTCCCTTGTATGAGGAGCTGTTTTCATGCTATTCACTTAT	363
Db	74	ACCGCTTTGGGATCCCTCCCTTGTATGAGGAGCTGTTTTCATGCTATTCACTTAT	15
QY	364	TAAATCTGCACT 377	
Db	14	TAAATCTGCACT 1	

RESULT 13
 LOCUS T69704/c 363 bp mRNA linear EST 07-MAR-1995
 DEFINITION yd3ia03.g1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
 IMAGE:66988 3', mRNA sequence.
 ACCESSION T69704
 VERSION T69704.1 GI:680852
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 363)
 AUTHORS Hillier, L., Clark, N., Dubque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lemon, G., Matra, M., Parsons, J.,
 Rifkin, L., Rohlfsing, T., Soares, M., Tan, F., Trevaekis, E., Waterston
 R., Williamson, A., Wohlmann, P. and Wilson, R.
 The Mashu-Merck EST Project
 Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 766
 High quality sequence stops: 341 Source: IMAGE Consortium, LNL. This
 clone is available royalty-free through LNL; contact the IMAGE
 Consortium (info@image.lnl.gov) for further information.
 Insert Length: 766 Std Error: 0.00
 Seq primer: -21ml3
 High quality sequence stop: 341.
 FEATURES
 source location/Qualifiers
 1..363
 /organism="Homo sapiens"
 /db_xref="GDB:463693"
 /db_xref="taxon:9606"
 /clone="IMAGE:66988"
 /clone_lib="Soares fetal liver spleen lNFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia)
 with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5', AACTGGAAGATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptor
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pTV73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonalido."
 BASE COUNT 83 a 83 c 87 g 107 t 3 others
 ORIGIN
 Query Match 65.7%; Score 329; DB 14; Length 363;
 Best Local Similarity 97.3%; Pred. No. 3.6e-96;
 Matches 354; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
 17 CCACACGACTTAGGTTTCTCTGTGAATGGGGCACTGAGAGACGACTAGCTGGATT 76
 Db 363 CCACACGACTTAGGTTTCTCTGTGAATGGGGCACTGAGAGACGACTAGCTGGATT 304
 Qy 77 TCGTAGGCTGACTAAGATCCCTAAGCCTAGCTAGGTGGGAAAGGTACACACATCCACTTTAAA 136
 Db 303 TCGTAGGCTGACTAAGATCCCTAAGCCTAGCTAGGTGGGAAAGGTACACACATCCACTTTAAA 244
 Qy 137 CACGGGCGTTGGCACTTAGCTCACAAGCTGACCAATCAAGAGAGCTCACTAAATGTCTAAAT 196
 Db 243 CACGGGCGTTGGCACTTAGCTCACAAGCTGACCAATCAAGAGAGCTCACTAAATGTCTAAAT 184
 Qy 197 AGCGAAGACAGAGAGTAAAGAAATGCAATCATCTATTGCTTAGAGACACAGCAGAG 256

Db 183 AGGCAAGACAGAGCTAAAGAAATAGCCATCATCTATTGCTTAGAGCACAGACAGAG 124

Qy 257 GGACATGATCGGGATATTAACCCAAAGTCTTGAGCCGGCAACGCCCTTTGGGT 316

Db 123 GGACATGATCGGGATATTAACCCAAAGTCTTGAGCCGGCAACGCCCTTTGGGT 66

Qy 317 CCCCTCCCTTTGATAGGAGCTCTGTTTTCATGCTATTGCTATTAAATCTTGCAAC 376

Db 65 CCCCTCCCTTTGATAGGAGCTCTGTTTTCATGCTATTGCTATTAAATCTTNAAAG 6

Qy 377 TGCA 380

Db 5 NACA 2

RESULT 14

AA860368/c 342 bp mRNA linear EST 31-DEC-1998

LOCUS a159c05.a1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1394600

DEFINITION 3', similar to contains PTR7.11 PTR5 repetitive element ;, mRNA sequence.

ACCESSION AA860368

VERSION AA860368.1 GI:2954363

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 342)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNI at: www-bio.illn.gov/bdtp/image/image.html
Insert Length: 1305 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 319.

FEATURES

SOURCE

1. .342

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1394600"

/clone_lib="Soares_testis_NHT"

/sex="male"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTACCAATGTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'], Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5 and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 75 a 78 c 82 g 107 t

ORIGIN

Query Match 65.2%; Score 326.6; DB 9; Length 342;

Best Local Similarity 98.8%; Pred. No. 2.2e-95;

Matches 329; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 48 GGGAGCTGAGAGACAGGACTAGCTGATTTCTTAGGCTGACTTAAGATCCCTTAAGCTTAG 107

Db 342 GGGAGCTGAGAGACAGGACTAGCTGATTTCTTAGGCTGACTTAAGATCCCTTAAGCTTAG 283

Qy 108 CTGGAGGTGACCAATCCACTTTTAAACAAGGGGCTTGGCACTTAGCTCAGACTGAC 167

Db 282 CTGGAGGTGACCAATCCACTTTTAAACAAGGGGCTTGGCACTTAGCTCAGACTGAC 223

Qy 168 CAATCAGAGACTCACTAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATAGCCAA 227

Db 222 CAATCAGAGACTCACTAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATAGCCAA 163

Qy 228 TCATCTATTGCTGAGAGACAGAGGAGGACATGATCGGGATATTAACCCAAAGTCTT 287

Db 162 TCATCTATTGCTGAGAGACAGAGGAGGACATGATCGGGATATTAACCCAAAGTCTT 103

Qy 288 CGAGCCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGATAGGAGAGCTGTTTCA 347

Db 102 CGAGCCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTTGATAGGAGAGCTGTTTCA 43

Qy 348 TGCTATTTCACCTATTATTAAATCTTGCAACTGCA 380

Db 42 TGCTATTTCACCTATTATTAAATCTTGCAACTGCA 10

RESULT 15

R27412/c 424 bp mRNA linear EST 24-APR-1995

LOCUS yH46d11.a1 Soares Placenta Mb2HP Homo sapiens cDNA clone

DEFINITION IMAGE:132789 3', mRNA sequence.

ACCESSION R27412

VERSION R27412.1 GI:783547

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 424)

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston,R., Williamson,A., Woldmann,P. and Wilson,R.

TITLE The Wash-Merck EST Project

JOURNAL Unpublished (1995)

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 695
High quality sequence stops: 367
Source: IMAGE Consortium, ILNI.
This clone is available royalty-free through ILNI; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 695 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 367.

FEATURES

SOURCE

1. .424

/organism="Homo sapiens"

/db_xref="GDB:538431"

/db_xref="taxon:9606"

/clone="IMAGE:132789"

/clone_lib="Soares placenta Mb2HP"

/sex="female"

/dev_stage="placenta obtained at birth (full term)"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGATTCGGCGCGAGATTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization.

BASE COUNT 92 a constructed by Bento Soares and M.Fatima Bernaldo. "
ORIGIN 96 c 115 g 121 t

Query Match 65.1%; Score 326.4; DB 14; Length 424;
Best Local Similarity 93.9%; Pred. No. 118e-95;
Matches 351; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

QY 4 TCGGCCAACCTGCCCAAGACGACTTAGGTTTCCTGTTGAGATGGGGGACTGAGAGACG 63
DB 373 TCGATCAACCTGCCCAAGACGACTTAGGTTTCCTGTTGAGATGGGGGACTGAGAGACG 314
QY 64 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTGACCAAC 123
DB 313 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTGACCTGC 254
QY 124 ATCCACCTTTAAACACGGGGGCTTGCACTAGCTGACCACTGACCAATCAGAGAGCTGAC 183
DB 253 ATCCACCTTTAAACACGGGGGCTTGCACTAGCTGACCACTGACCAATCAGAGAGCTGAC 196
QY 184 TAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 243
DB 195 TAAATGCTAATTAGGCAAGAAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 136
QY 244 AGCAGAGAGAGGAGCAATGATGGGATATAACCTCAAGCTTGAGCCGCAAGGCA 303
DB 135 AGCAGAGAGAGGAGGCAAGATCGGATATAACCTCAAGCTTGAGCCGCAAGGCA 76
QY 304 ACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTGTTTCATGCTATTGACTTAT 363
DB 75 AACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTGTGTTTCATGCTATTGACTTAT 16
QY 364 TAAATCTGCAACT 377
DB 15 TAAATCTGCAACT 2

Search completed: April 19, 2003, 14:46:37
Job time : 1271.51 secs

JOURNAL Patent: WO 9367395-A 3 29-DEC-1993;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES Location/Qualifiers
SOURCE 1..10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2455 t
ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 10499;
Best Local Similarity 100.0%; Pred. No. 6,2e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTCTGGCCAACTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 60
DB 9500 GCTCTGGCCAACTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 9559
OY 61 CAGACTAGCTGATTTCTTGAAGTAAATCTTAACTTGAAGCTGGAAGGTGAC 120
DB 9560 CAGACTAGCTGATTTCTTGAAGTAAATCTTAACTTGAAGCTGGAAGGTGAC 9619
OY 121 CACATCCACTTTAAACAGAGGGCTTGCATTAGTCAACCTGACCAATGAGAGCT 180
DB 9620 CACATCCACTTTAAACAGAGGGCTTGCATTAGTCAACCTGACCAATGAGAGCT 9679
OY 181 CACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCT 240
DB 9680 CACTAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCT 9739
OY 241 GAGACACAGCAGAGAGGCAATGATCGGATATTAACCAAGCTTTCGAGCCGCAACG 300
DB 9740 GAGACACAGCAGAGAGGCAATGATCGGATATTAACCAAGCTTTCGAGCCGCAACG 9799
OY 301 GCAACCCCTTTGGGTCCTCCCTTTGATGGAGTCTGTGTTTCAAGCTATTTCACTC 360
DB 9800 GCAACCCCTTTGGGTCCTCCCTTTGATGGAGTCTGTGTTTCAAGCTATTTCACTC 9859
OY 361 TATTAAATCTTCACTGACTCTTCTGTCATGTTCTTACGGCTTGAAGCTGAGCTT 420
DB 9860 TATTAAATCTTCACTGACTCTTCTGTCATGTTCTTACGGCTTGAAGCTGAGCTT 9919
OY 421 CGCTGCGCATCCACCACTGCTGTTTGGCCGCAACCGGAGACCCGCGCTGATCCCATCC 480
DB 9920 CGCTGCGCATCCACCACTGCTGTTTGGCCGCAACCGGAGACCCGCGCTGATCCCATCC 9979
OY 481 TCTGATCATGCAAGGTGTC 501
DB 9980 TCTGATCATGCAAGGTGTC 10000

RESULT 2
AX329572 56093 bp DNA linear PAT 09-JAN-2002
LOCUS AX329572
DEFINITION Sequence 81 from Patent WO0194629.
ACCESSION AX329572
VERSION AX329572.1 GI:18102550
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ehner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES location/Qualifiers
SOURCE 1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 99.4%; Score 498; DB 6; Length 56093;
Best Local Similarity 100.0%; Pred. No. 7,2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TCGGCCAACTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 63
DB 37504 TCGGCCAACTCCCAACAGCACTTGTGTTTCTGTGAGATGGGGAGCTAGAGA 37563
OY 64 GACTAGCTGATTTCTTGAAGTAAATCTTAACTTGAAGCTGGAAGGTGAC 123
DB 37564 GACTAGCTGATTTCTTGAAGTAAATCTTAACTTGAAGCTGGAAGGTGAC 37623
OY 124 ATTCACCTTTAAACAGAGGGCTTGAAGTAAATCTTAACTTGAAGCTGGAAGGTGAC 183
DB 37624 ATTCACCTTTAAACAGAGGGCTTGAAGTAAATCTTAACTTGAAGCTGGAAGGTGAC 37683
OY 184 TAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 243
DB 37684 TAAATGCTAATTAAGCAAGAGAGGTAAAGAAATAGCAATCATCTATTGCTGAG 37743
OY 244 AGCAGCAGAGAGGCAATGATCGGATATTAACCAAGCTTTCGAGCCGCAACGCA 303
DB 37744 AGCAGCAGAGAGGCAATGATCGGATATTAACCAAGCTTTCGAGCCGCAACGCA 37803
OY 304 ACCCCCTTTGGGTCCTCCCTTTGATGGAGTCTGTGTTTCAAGCTATTTCACTCTAT 363
DB 37804 ACCCCCTTTGGGTCCTCCCTTTGATGGAGTCTGTGTTTCAAGCTATTTCACTCTAT 37863
OY 364 TAAATCTTGAACCTGACTCTTCTGTCATGTTCTTACGGCTTGAAGCTGAGCTTTCG 423
DB 37864 TAAATCTTGAACCTGACTCTTCTGTCATGTTCTTACGGCTTGAAGCTGAGCTTTCG 37923
OY 424 TCGCATCCACCACTGCTGTTTGGCCGCAACCGGAGACCCGCGCTGATCCCATCCCT 483
DB 37924 TCGCATCCACCACTGCTGTTTGGCCGCAACCGGAGACCCGCGCTGATCCCATCCCT 37983
OY 484 GGATCATGCAAGGTGTC 501
DB 37984 GGATCATGCAAGGTGTC 38001

RESULT 3
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996
LOCUS HSAC000064
DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.
ACCESSION AC000064
VERSION AC000064.1 GI:1669369
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 56093)
AUTHORS Paulley, A.
TITLE The sequence of H. sapiens BAC clone RG083M05
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 56093)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-1996)
COMMENT Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: saplens@wustl.edu

NOTICE: This sequence may not represent the entire insert of this
section. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBET0

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SWS1725.

FEATURES

SOURCE

location/Qualifiers

1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"

/clone_id="CTTB-978SK-B"
complement(838..1131)
/rpt_family="ALU"

repeat_region

/rpt_family="ALU"

gene

/gene="WUGSC:H_RG083M05.1"

CDS

join(<1360..1503,4181..4370,4587..4774,6422..6556,
9483..9547,11631..11773,11864..12021,13131..13296,
14885..14988,16349..16546,16837..16971)
/gene="WUGSC:H_RG083M05.1"
/note="ATPase; strong similarity to peroxisome
biosynthesis protein Pasi (PID:g1172019); coded for by
human CDNA C04279 (NID:g146530)"

/codon_start=1
/protein_id="AAB46346.1"

/db_xref="GI:1669371"

/translation="KRLENIQKLEVAFAEAMVQPSVLLDDLLIAGLPAVEHEH
SPDAORCEILCNVKNKLDGDKFTDLOHAKKEGCVADFTLVDRALHSL
SRQISTRREKVLTTDPRKALRGTRPLSLSSVNHKRPDGMKIGLHVRQLMD
TTQIPAKVCLKEKYPPELFAVLPTRQRTGILLYGPDTGKTLGAVIARESRNFISV
KGPBLISKYIGASQAVADIFIRQAAPCILFDEFESIAPRGHDTVTDVVNO
LITQIDVGEIGAGVAVLAATSRPLIDPALRPGLDKCVCPDPDVTTSYLSKTKQ
QMLHSFLVSRLEINLVLSDSLPLADVDVLOHVASVTSFGADLKALLYNQLBALHG
MLLSKMSRIIPDESKENMYRLYFGSSVSEELGNGSSDLSGCLSAPESTMQDLRGVP
GKDDLFSGPPVLRTRASQGCCELTQEQKQDQRADISITIKGYSQSGEDSMQPGPI
KTRLAISQSHLMTALGTRPSISEDMKNFAEL"

repeat_region

complement(4948..5130)

repeat_region

/rpt_family="ALU"

repeat_region

complement(6581..7133)

repeat_region

/rpt_family="ALU"

repeat_region

complement(8186..8472)

repeat_region

/rpt_family="ALU"

repeat_region

8473..8625

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 3' EST H75782 (NID:g1049794), bases
287-444"

repeat_region

8841..9161

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H75921 (NID:g1050050), bases
21-348"

repeat_region

9481..9547

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST N22627 (NID:g1130501), bases
276-343"

repeat_region

complement(12612..12907)

repeat_region

/rpt_family="ALU"

repeat_region

13670..13793

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H41382 (NID:g17434), bases
143-266"

repeat_region

13794..13877

repeat_region

/rpt_family="ALU"

repeat_region

13878..13906

repeat_region

/gene="WUGSC:H_RG083M05.1"

repeat_region

/note="match to human 5' EST H41382 (NID:g917434), bases
30-58"

repeat_region

13907..14104

repeat_region

/rpt_family="ALU"

repeat_region

complement(14110..14137)

repeat_region

/rpt_family="ALU"

repeat_region

complement(15618..15907)

repeat_region

/rpt_family="ALU"

repeat_region

17227..17522

repeat_region

/rpt_family="ALU"

repeat_region

18667..19235

repeat_region

/note="match to human fetal brain 5' EST D61494
(NID:g970409), bases 1-255, and to human 3' EST R07476
(NID:g759399)"

repeat_region

19550..19670

repeat_region

/rpt_family="ALU"

repeat_region

21507..37303

repeat_region

/note="similarity to various SS-RNA virus polypeptides;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:g842257)"

repeat_region

37316..37489

repeat_region

/note="Grail prediction, score = 80"

repeat_region

/evidence=not experimental

repeat_region

complement(38938..39224)

repeat_region

/rpt_family="ALU"

repeat_region

39225..39707

repeat_region

/note="match to multiple human ESTs, see N30113
(NID:g1148633)"

repeat_region

39800..40085

repeat_region

/rpt_family="ALU"

repeat_region

complement(40247..40538)

repeat_region

/rpt_family="ALU"

repeat_region

complement(40632..40924)

repeat_region

/rpt_family="ALU"

repeat_region

complement(42283..42891)

repeat_region

/rpt_family="ALU"

repeat_region

complement(44574..45613)

repeat_region

/rpt_family="ALU"

repeat_region

complement(45614..45737)

repeat_region

/note="match to human 3' EST H48898 (NID:g988738), bases
129-333"

repeat_region

complement(46107..47026)

repeat_region

/note="match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and W78831
(NID:g273146)"

repeat_region

complement(47027..47318)

repeat_region

/rpt_family="ALU"

repeat_region

complement(47365..47782)

repeat_region

/note="match to multiple human ESTs, see W37495
(NID:g1319089)"

repeat_region

47898..48115

repeat_region

/note="match to human 5' EST H62306 (NID:g1015138), bases
93-368"

repeat_region

complement(48116..48405)

repeat_region

/rpt_family="ALU"

repeat_region

complement(48406..48584)

repeat_region

/note="match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:g765806)"

repeat_region

complement(48787..49405)

repeat_region

/rpt_family="ALU"

repeat_region

complement(49406..49534)

repeat_region

/note="match to human 3' EST R65794 (NID:g838432), bases
309-440"

```

repeat_region      complement(49638..49672)
                    /rpt_family="ALU"
misc_feature        complement(49674..49896)
                    /note="match to human 3' EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148458), sequences are from opposite
                    ends of the same clone"
gene                complement(49698..51806)
                    /gene="WUGSC:H.RG083M05.2"
                    complement(join(49698..49888,51575..51806))
                    /gene="WUGSC:H.RG083M05.2"
                    /note="coded for by human cDNA M37389 (NID:g119205),
                    R65891 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
                    /codon_start=1
                    /protein_id="AAB46345.1"
                    /db_xref="GI:1669370"
                    /translation="MFYFVFOCGI::FCPPGVVYQIGNVSVIDEQDKPYVAGIRP
                    IODICEKSAALTWLITPLSSPRDIFDPASTITIGBEDLPFKMEYLEFVCHAPSEYFK
                    SRSSPFPVTPRPEKGIWTVGP:PALTIKESVANHL"
exon                complement(51576..51758)
                    /gene="WUGSC:H.RG083M05.2"
                    /note="Grail prediction, score = 86"
                    /evidence=not_experimental
repeat_region      complement(52052..52329)
                    /rpt_family="h1"
                    /note="match to human EST M79192 (NID:g273505) base 2-289"
misc_feature        55557..55843
Query Match        99.4%; Score 498; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 7.2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

4 TCGGCAACCTCCCAACGACCTAGGTTTCCTGTTGAGATGGGGGACGACGAGACG 63
 Db 37504 TCGGCAACCTCCCAACGACCTAGGTTTCCTGTTGAGATGGGGGACGACGAGACG 37563
 Qy 64 GACTAGCTGAGATTCTCTAGGCTGACTAAGAAATCCCTAGCTGAGAGGTGACAC 123
 Db 37564 GACTAGCTGAGATTCTCTAGGCTGACTAAGAAATCCCTAGCTGAGAGGTGACAC 37623
 Qy 124 ATCCACCTTTAAACAGGGGGCTTGCACCTTAGCTCA(C)ACCTAGCCATATAGAGACTCAC 183
 Db 37624 ATCCACCTTTAAACAGGGGGCTTGCACCTTAGCTCA(C)ACCTAGCCATATAGAGACTCAC 37683
 Qy 184 TAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATAGCAATCATCTATTTGCTGAG 243
 Db 37684 TAAATGCTAATTAGGCAAGAAGAGAGGTAAAGAAATAGCAATCATCTATTTGCTGAG 37743
 Qy 244 AGCAGCAGCAGAGGAGCAATGATCGGAGATATAACCCCAAGTCTTGCAGCCGCAACGGCA 303
 Db 37744 AGCAGCAGCAGAGGAGCAATGATCGGAGATATAACCCCAAGTCTTGCAGCCGCAACGGCA 37803
 Qy 304 ACCCCCTTTGGGTCCCTCTTGTATGGAGCTCTATTTCATGCTATTTCACCTCTAT 363
 Db 37804 ACCCCCTTTGGGTCCCTCTTGTATGGAGCTCTATTTCATGCTATTTCACCTCTAT 37863
 Qy 364 TAAATCTTGCAACTGACCTCTTGTGCTCATGTTCTTACGGCTTGAGCTGAGCTTTGCC 423
 Db 37864 TAAATCTTGCAACTGACCTCTTGTGCTCATGTTCTTACGGCTTGAGCTGAGCTTTGCC 37923
 Qy 424 TCGCATCCACCACTGCTGTTTGCAGCAACCGCAGAC(C)GCGCTGACTCCCATCCCTCT 483
 Db 37924 TCGCATCCACCACTGCTGTTTGCAGCAACCGCAGAC(C)GCGCTGACTCCCATCCCTCT 37983
 Qy 484 GGATCATGCAAGGCTGCC 501
 Db 37984 GGATCATGCAAGGCTGCC 38001

RESULT 4
 AC007566/c AC007566 149194 bp DNA linear PRI 01-MAR-2002
 LOCUS Homo sapiens BAC clone CTB-1005 from 7q21-7q22, complete sequence.
 DEFINITION

ACCESSION AC007566
 VERSION AC007566.2 GI:11181861
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE AUTHORS Sulston, J.E. and Waterston, R.
 JOURNAL Toward a complete human genome sequence
 MEDLINE Genome Res. 8 (11), 1097-1108 (1998)
 PUBMED 99063792
 REFERENCE PUBMED 9847074
 REFERENCE Du, Z.
 TITLE The sequence of Homo sapiens BAC clone CTB-1005
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 149194)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAY-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 149194)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (02-OCT-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 5 (bases 1 to 149194)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 6 (bases 1 to 149194)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 7 (bases 1 to 149194)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (06-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 8 (bases 1 to 149194)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Nov 16, 2000 this sequence version replaced gt:4835815.
 COMMENT ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_RG010605

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send <mailto:malco@egenomhgrl.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.reagen.com/>).

VECTOR: pBelOBAc11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5, actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

SOURCE

Location/Qualifiers
1. .149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_lib="CTB-978SK-B"
1. .1634
/rpt_family="L1"
/note="match to EST BG752883 (NID:g14063536)"
2248. .2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
2248. .2387
/note="match to EST AM579261 (NID:g7254310)"
2248. .2387
/note="match to EST BG766882 (NID:g14077535)"
2248. .2387
/note="match to EST BI160365 (NID:g14620366)"
2248. .2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
2248. .2287
/note="match to EST AV686676 (NID:g10288539)"
2253. .2387
/note="match to EST AU123510 (NID:g10948226)"
2344. .2387
/note="match to EST AV686676 (NID:g10288539)"
2696. .3066
/rpt_family="MALR"
3108. .3392
/rpt_family="A1u"
3540. .3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
3540. .3628
/note="similar to Mus musculus EST BE253526 (NID:g8946272)"
3542. .3632
/note="match to EST AV686676 (NID:g10288539)"
3542. .3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
3542. .3628
/note="match to EST AU123510 (NID:g10948226)"
3542. .3628
/note="match to EST BG766882 (NID:g14077535)"

misc_feature 3542. .3628
/note="match to EST BI160365 (NID:g14620366)"
misc_feature 3542. .3609
/note="match to EST AM579261 (NID:g7254310)"
misc_feature 3562. .3628
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3598. .3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714. .3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717. .3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719. .3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719. .3785
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature 3719. .3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719. .3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3719. .3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719. .3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719. .3785
/note="match to EST BI160365 (NID:g14620366)"
misc_feature 3719. .3785
/note="similar to Mus musculus EST BE253526 (NID:g8946272)"
misc_feature 3860. .3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730. .4827
/rpt_family="L2"
repeat_region 4834. .4925
/rpt_family="L1"
misc_feature 5715. .5860
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature 5715. .5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 5715. .5860
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 5715. .5860
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 5715. .5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 5715. .5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 5715. .5860
/note="match to EST BI160365 (NID:g14620366)"
misc_feature 5715. .5860
/note="similar to Mus musculus EST BE994936

Query Match 99.4%; Score 498; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 8, 2e-150;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 TCGGCAACCTCCCAACAGCACTTAGCTTTCTCTGTGAGATGGGGACTGAGAGACG 63
Db 84500 TCGGCAACCTCCCAACAGCACTTAGCTTTCTCTGTGAGATGGGGACTGAGAGACG 84441

Qy 64 GACTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAGGCTAGCTGGGAAGTGACAC 123
Db 84440 GACTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAGGCTAGCTGGGAAGTGACAC 84381

Qy 124 ATCCACCTTTAAACAGGGGCTTGCAACTTAGCTACACCTGACCAATCAGAGAGCTTAC 183
Db 84380 ATCCACCTTTAAACAGGGGCTTGCAACTTAGCTACACCTGACCAATCAGAGAGCTTAC 84321

Qy 184 TAAATGCTAATTGGCAAGACGAGGTTAAAGAAATAGCCATCATCTTTTCCTGAG 243
Db 84320 TAAATGCTAATTGGCAAGACGAGGTTAAAGAAATAGCCATCATCTTTTCCTGAG 84261

QY 244 AGCAGCAGAGAGGAGCAATGATCGGGATATTAAC CCAAGTCTTGAGCCGGCAACGCGA 303
DB 84260 AGCAGCAGAGAGGAGCAATGATCGGGATATTAAC CCAAGTCTTGAGCCGGCAACGCGA 84201
QY 304 ACCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTCTGTTTTCAGTATTTCACTCAT 363
DB 84200 ACCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTCTGTTTTCAGTATTTCACTCAT 84141
QY 364 TAAATCTTGCAACTGACATCTTCTGATCATGTTT TTAACGGCTTGAGCTGAGCTTTGCG 423
DB 84140 TAAATCTTGCAACTGACATCTTCTGATCATGTTT TTAACGGCTTGAGCTGAGCTTTGCG 84081
QY 424 TCGCCATCCACCATGCTGTTTTCGCGCAACGCGAACCCTGCTGACTGCCATCCCTCT 483
DB 84080 TCGCCATCCACCATGCTGTTTTCGCGCAACGCGAACCCTGCTGACTGCCATCCCTCT 84021
QY 484 GGATCATGAGGTGTCC 501
DB 84020 GGATCATGAGGTGTCC 84003

RESULT 5
AX007998
LOCUS AX007998 711 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 21 from Patent WO9967395.
ACCESSION AX007998
VERSION AX007998.1 GI:995695
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 711)
AUTHORS Perin,J.P., Rieger,P. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL Patent: WO 9967395-A 21 29-DEC-1999;
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES
Location/Qualifiers
source 1..711
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 171 a 209 c 160 g 171 t
ORIGIN

Query Match 89.8%; Score 450; LB 6; Length 711;
Best Local Similarity 100.0%; Pred. No. 1.6e-134; Indels 0; Gaps 0;
Matches 450; Conservative 0; Mismatches 0;

QY 52 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCTGACCTAAGCTAGCTGG 111
DB 1 ACTGAGAGCAGAGCTAGCTGATTTCTTGAAGCTGACCTAAGCTAGCTGG 60
QY 112 GAAGTGACACATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCACCTGACCAAT 171
DB 61 GAAGTGACACATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCACCTGACCAAT 120
QY 172 CAGAGAGCTCACTAAATGCTAATTAAGCAAGAGAGGTAAGAAATTAACCAATCT 231
DB 121 CAGAGAGCTCACTAAATGCTAATTAAGCAAGAGAGGTAAGAAATTAACCAATCT 180
QY 232 CTAATGCTGAGAGCAGCAGAGAGGAGCAATGATCGATATTAACCAATCTTTCAG 291
DB 181 CTAATGCTGAGAGCAGCAGAGAGGAGCAATGATCGATATTAACCAATCTTTCAG 240
QY 292 CCGGCAACGGCAACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTCTGTTTCACTGCT 351
DB 241 CCGGCAACGGCAACCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTCTGTTTCACTGCT 300
QY 352 ATTGACTCTATTAAATCTTTCGACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGA 411
DB 301 ATTGACTCTATTAAATCTTTCGACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGA 360

QY 412 CTGAGCTTTCGCTCGCATTCACCATGCTGTTTTCGCGCAACGCGAACCCTGCTGAC 471
DB 361 CTGAGCTTTCGCTCGCATTCACCATGCTGTTTTCGCGCAACGCGAACCCTGCTGAC 420
QY 472 TCCCATCCCTTGGATCATGAGAGGTGTCC 501
DB 421 TCCCATCCCTTGGATCATGAGAGGTGTCC 450

RESULT 6
AP001538/c 174019 bp DNA linear PRI 25-MAR-2000
LOCUS AP001538
DEFINITION Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
L156-APP region, complete sequence.
ACCESSION AP001538
VERSION AP001538.1 GI:7328982
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B853K11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 174019)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 174019)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (23-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
Kilasato Univ., 1-15-1 Kilasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsr.riken.go.jp, Tel:01-42-778-9923,
Fax:01-42-778-9924)
FEATURES
Location/Qualifiers
source 1..174019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1-q21.2"
/clone="B853K11"

BASE COUNT 53303 a 31416 c 32248 g 57052 t
ORIGIN

Query Match 86.6%; Score 434; DB 9; Length 174019;
Best Local Similarity 92.0%; Pred. No. 4.9e-129; Indels 0; Gaps 0;
Matches 458; Conservative 0; Mismatches 40;

QY 4 TCGGCAACCTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGAGCTGAGAGACAG 63
DB 36468 TTGGCCAACTCCCAACAGCAGCTTAGGTTTCTGTTGAGATGGGAGCTGAGAGACAG 36409
QY 64 GACTAGCTGATTTCTTGAAGCTGATCCCTTAAGCTAGCTGGAGAGTGACCA 123
DB 36408 GACTAGCTGATTTCTTGAAGCTGATCCCTTAAGCTAGCTGGAGAGTGACCA 36349
QY 124 ATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCACCTGACCAATCAAGAGCTTAC 183
DB 36348 ATCCACCTTTAAACAGGGGCTTGCACCTTAGCTCACCTGACCAATCAAGAGCTTAC 36289
QY 184 TAAATGCTAATTAAGCAAGAGAGAGCTAAGAAATAGCAATCATCTATTGCTGAG 243
DB 36288 TAAATGCTAATTAAGCAAGAGAGAGCTAAGAAATAGCAATCATCTATTGCTGAG 36229
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCCAAGTCTTTCGAGCCGGCAACGCGCA 303
DB 36228 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCCAAGTCTTTCGAGCCGGCAACGCGCA 36169
QY 304 ACCCCCTTTGGGTCCCTTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTCTCAT 363

Db 36168 ACCCCCTTGGGTCCCTGCTTGTGATGGAGCTCTGTTTACTCTATTTACTCTAC 36109
Qy 364 TAAATCTGCAACTGACCTCTTCTGTCATGTTTCTTACGAGCTTGTAGCTGTTTGC 423
Db 36108 TAAATCTGCAACTGACCTCTTCTGTCATGTTTCTTACGAGCTTGTAGCTGTTTGC 36049
Qy 424 TCGGCATCCACCACTGCTGTTTGGCGCAGACCCGCGCTGACTCCATCCCTCT 483
Db 36048 TCGCTTCCACCACTGCTGTTTGGCGCAGACCTGCACTGACTCCATCCCTCTCC 35989
Qy 484 GGATTCATCAGAGGTCTCC 501
Db 35988 GGATTCGGCAGAGCTGTCC 35971

RESULT 7
LOCUS AP001674/c 340000 bp DNA linear PRI 10-MAY-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 18/105.
ACCESSION AP001674 AL163219 BA000005
VERSION AP001674.1 GI:7768666
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 340000)
AUTHORS Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Paterson,D., Reichwald,K., Rump,A., Schilhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstrem,G.,
Hornischler,K., Barand,P., Scharte,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hemmi,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaepo,M.L.
The DNA sequence of human chromosome 21
Nature 405 (6784), 311-319 (2000)
2 (bases 1 to 340000)
20289799

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

COMMENT
* e-mail: hattori@cec.riken.go.jp
* URL: http://hgp.cec.riken.go.jp/
and
Beutenbergstrasse 11, D-07745 Jena, Germany, *

FEATURES
source
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
1. 6149
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R749F18, 5' partial"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AP001537"
1. 165231
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R853K11, 5' partial"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AP001538"
136796. 315056
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R845I12"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AF245226"
156433. 277809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="P646P19"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AL157359"
207594. 250588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="Q3D19"
/clone_1lb="AL121NC02-Q Coamid library"
/note="Accession No. AL049753"
276958. 301481
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R755I19"
/clone_1lb="RPCL-11 BAC library"
/note="Accession No. AP001345"
299634. 340000
/organism="Homo sapiens"

source	/db_xref="taxon:9606"
	/chromosome="21"
	/map="21q21.1"
	/clone="R28M9, 3' partial"
	/clone_1lb="RPC1-1 BAC library"
	/note="Accession No. AP000431"
	311751. .>340000
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="21"
repeat_region	/map="21q21.1"
	/clone="103E2331, 3' partial"
	/clone_1lb="HRCPC103 Combid library"
	/note="Accession No. AL332993"
	15. .444
	/note="L1MA4A"
	/rpc_family="LINE/L1"
	/rpc_type=DISPERSED
	555. .3531
	/note="L1MA4A"
repeat_region	/rpc_family="LINE/L1"
	/rpc_type=DISPERSED
	3770. .3876
	/note="L2"
	/rpc_family="LINE/L2"
	/rpc_type=DISPERSED
	complement(5645. .6178)
	/note="L1MA2"
	/rpc_family="LINE/L1"
	/rpc_type=DISPERSED
repeat_region	6181. .6534
	/note="THE1B"
	/rpc_family="LTR/MaLR"
	/rpc_type=DISPERSED
	8021. .8370
	/note="THE1B"
	/rpc_family="LTR/MaLR"
	/rpc_type=DISPERSED
	complement(8373. .10032)
	/note="L1MA2"
repeat_region	/rpc_family="LINE/L1"
	/rpc_type=DISPERSED
	10213. .10520
	/note="Aluub"
	/rpc_family="SINE/Alu"
	/rpc_type=DISPERSED
	15810. .15851
	/note="(A)n"
	/rpc_family="Simple_repeat"
	/rpc_type=TANDEM
repeat_region	complement(16628. .16697)
	/note="MER53"
	/rpc_family="DNA"
	/rpc_type=DISPERSED
	16638. .16715
	/note="MER53"
	/rpc_family="DNA"
	/rpc_type=DISPERSED
	18847. .19585
	/note="L2"
repeat_region	/rpc_family="LINE/L2"
	/rpc_type=DISPERSED
	19795. .19917
	/note="(TA)n"
	/rpc_family="Simple_repeat"
	/rpc_type=TANDEM
	20023. .20166
	/note="(TA)n"
	/rpc_family="Simple_repeat"
	/rpc_type=TANDEM
repeat_region	20285. .20448
	86.64; Score 434; DE 9; Length 340000

Best Local Similarity	92.0%	Pred	No. 5'-129;		
Matches	458;	Conservative	0;	Mismatches	40;
				Indels	0;
				Gaps	0;
QY	4	TGGGCAACCTCCCCAACAGACACTTAAAGTTTTCTGTGTGAGATGGGGAGCTGAGAGACAG	63		
Db	27679	TTGGCAACCTCCCCAACAGACACTTGGGTTTTTCTGTGTGAGAGGGGGTACGAGAGACAG	27620		
QY	64	GACTAGCTGSAATTTCTTAGGCTGACTAAGAAATCCCTAAGCTAGCTGGGAAGTGACAC	123		
Db	27619	GACTAGCTGSAATTTCTTAGGCTGACTAAGAAATCCCTAAGCTAGCTGGGAAGTGACCGC	27560		
QY	124	ATCCACTTTTAAACAGGGGGCTTGCAATTAGCTCACACCTGACCAATCAGAGGCTCAC	183		
Db	27559	ATCCACTTTTAAACATGGGGGCTCGCACTTGACTCACCCAACCAAGTCAGAGACTCAC	27500		
QY	184	TAAATGCTAATTAGGCAAGCAGAGAGTAAAGAAATAGCCAATCATCTAATTGGCTAG	243		
Db	27499	TAAATGCTAATTAGGCAAAACAGAGATTAAGAAATAGCCAATCATCTAATTGGCTAG	27440		
QY	244	AGCAGACAGAGAGGAGACATGATCGAGATATATAACCCAGTCTTCGAGCCGGCAACGCA	303		
Db	27439	AGCAGACAGAGAGGAGACAGAGATCAGATGTAAACCCAGGACTTAACTGGCAACGGCA	27380		
QY	304	ACCCTCTTGGGTCCCTCCCTCCCTTGTATGGAGCTCTGTTCATGACTAATTCATCTAT	363		
Db	27379	ACCCTCTTGGGTCCCTCCCTCTTGTATGGAGCTCTGTTCATCTAATTCATCTATAC	27320		
QY	364	TAAATCTTGCAATGCACTCTTGTGTGCATAGTTTCTTAAGCGCTTGAGCTGACGTTTGGC	423		
Db	27319	TAAATCTTGCAATGCACTCTTGTGTGCATAGTTTGTTAAGCGCTTGAGCTGAGCTTGGC	27260		
QY	424	TGCGCATCCACACATGCTGTTTGCCGCCACAGCCAGACCCGCGCTGACTCCCATCCCTCT	483		
Db	27259	TGCGTTTCACACATGCTGTTTGCCGCTGTGCAGACCTGCCACTGACTCCCATCCCTCC	27200		
QY	484	GGATCATCAGAGGCTGCC	501		
Db	27199	GGATCCGACAGGCTGCC	27182		

RESULT 8	AC093531/c	AC093531	163803 bp	DNA	linear	PRI 16-NOV-2001
LOCUS						
DEFINITION		Homo sapiens chromosome 5 clone RP11-405L7,			complete sequence.	
ACCESSION		AC093531				
VERSION		AC093531.2			GI:16945981	
KEYWORDS		HTG.				
SOURCE		Homo sapiens.				
ORGANISM		Homo sapiens				
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE		1 (bases 1 to 163803)				
AUTHORS		DOE Joint Genome Institute and Stanford Human Genome Center.				
TITLE		Direct Submission				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 163803)				
AUTHORS		DOE Joint Genome Institute.				
TITLE		Direct Submission				
JOURNAL		Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint				
REFERENCE		Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
AUTHORS		3 (bases 1 to 163803)				
TITLE		DOE Joint Genome Institute and Stanford Human Genome Center.				
JOURNAL		Direct Submission				
		Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell				
		Drive, Walnut Creek, CA 94598, USA				
COMMENT		On Nov 16, 2001 this sequence version replaced gi:15383820.				
		Draft Sequence Produced by DOE Joint Genome Institute				
		www.jgi.doe.gov				
		Finishing completed at Stanford Human Genome Center				
		www.sbgc.stanford.edu				
		Quality: Phrap Quality >=40 99.8% of Sequence;				
		Estimated Total Number of Errors is 0.3.				
FEATURES		Location/Qualifiers				

FEATURES

Location/Qualifiers

source 1.163803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"

BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN

Query Match 86.1%; Score 431.2; DB 9; Length 163803;
Best Local Similarity 91.4%; Pred. No. 4e-128;
Matches 457; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2 TCTGGCCCAACCTCCCAACGCACTTGAAGTTCCTTGAAGTGGGGAGCTGAGAGAC 61
Db 117108 TGTGAGCCCAACCTCCCAACGCACTTGAAGTTCCTTGAAGTGGGGAGCTGAGAGAC 117049

QY 62 AGGACTGAGTGGATTTCTAGGCTGACCTAAGATCCCTAAGCTTGAAGTGGAGAGTGC 121
Db 117048 AGGACTGAGTGGATTTCTAGGCTGACCTAAGATCCCTAAGCTTGAAGTGGAGAGTGC 116989

QY 122 ACATCCACCTTTAAACAAGGGGCTTGCACCTAGCTCACCTGACCAATCAGAGAGCTC 181
Db 116988 GCATCCACCTTTAAACAAGGGGCTTGCACCTAGCTCACCTGACCAATCAGAGAGCTC 116929

QY 182 ACTAAATGCTAATTAAGCAAGAGGTAAGAAATAGCAATCATATATTCCTG 241
Db 116928 ACTAAATGCTAATTAAGCAAGAGGTAAGAAATAGCAATCATATATTCCTG 116869

QY 242 AGAGCAGCAGCAGGAGGACATGATCGGATATTAACCAAGCTTTCAGAGCCGCAAGG 301
Db 116868 AGAGCAGCAGCAGGAGGACATGATCGGATATTAACCAAGCTTTCAGAGCCGCAAGG 116809

QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTGTTTCATGCTATTTCACTCT 361
Db 116808 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTGTTTCATGCTATTTCACTCT 116749

QY 362 ATTAATCTTGCAACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGAAGCTTTC 421
Db 116748 ATTAATCTTGCAACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGAAGCTTTC 116689

QY 422 GCTGCCATCCACACGCTGCTTTTGGCCGACCCGACCCGCGCTGACCTCCATCCCT 481
Db 116688 GCTGCCATCCACACGCTGCTTTTGGCCGACCCGACCCGCGCTGACCTCCATCCCT 116629

QY 482 CTGATCATGACGAGGTGTC 501
Db 116628 TTGATTCACGACGAGTGTCC 116609

RESULT 9
AF127226 839 bp DNA linear PRI 17-JUN-1999
LOCUS Homo sapiens Human endogenous retrovirus W, 3' long terminal
DEFINITION repeat, partial sequence.
ACCESSION AF127226
VERSION AF127226.1 GI:5081468
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 839)
AUTHORS Ounanian-Paraz, A., Komurian-Pradel, F., Otc, C., Rajorharison, A. and
Perron, H.
TITLE MSRV-related HERV-W elements in human DNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 839)
AUTHORS Ounanian-Paraz, A., Komurian-Pradel, F., Otc, C., Rajorharison, A. and
Perron, H.
TITLE Direct Submission
JOURNAL Submitted (10-FEB-1999) UMR103 CNRS-Biomerieux, Biomerieux, 46,
Allee D' Italie, Lyon 69007, France
FEATURES
Location/Qualifiers

source 1.839
/organism="Homo sapiens"
/db_xref="taxon:9606"
LTR <1..>839
/note="3', long terminal repeat; U3, R and U4 regions"

repeat_region 1.839
/note="MSRV-related"
/rpt_family="Human endogenous retrovirus W"
/rpt_type="dispersed"

BASE COUNT 205 a 233 c 192 g 206 t 3 others
ORIGIN

Query Match 85.0%; Score 426; DB 9; Length 839;
Best Local Similarity 90.6%; Pred. No. 9.9e-127;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCCAACCTCCCAACGCACTTGAAGTTCCTTGAAGTGGGGAGCTGAGAGAC 61
Db 4 TGTGAGCCCAACCTCCCAACGCACTTGAAGTTCCTTGAAGTGGGGAGCTGAGAGAC 63

QY 62 AGGACTGAGTGGATTTCTAGGCTGACCTAAGATCCCTAAGCTTGAAGTGGAGAGTGC 121
Db 64 AGGACTGAGTGGATTTCTAGGCTGACCTAAGATCCCTAAGCTTGAAGTGGAGAGTGC 123

QY 122 ACATCCACCTTTAAACAAGGGGCTTGCACCTAGCTCACCTGACCAATCAGAGAGCTC 181
Db 124 GCATCCACCTTTAAACAAGGGGCTTGCACCTAGCTCACCTGACCAATCAGAGAGCTC 183

QY 182 ACTAAATGCTAATTAAGCAAGAGGTAAGAAATAGCAATCATATATTCCTG 241
Db 184 ACTAAATGCTAATTAAGCAAGAGGTAAGAAATAGCAATCATATATTCCTG 243

QY 242 AGAGCAGCAGCAGGAGGACATGATCGGATATTAACCAAGCTTTCAGAGCCGCAAGG 301
Db 244 AGAGCAGCAGCAGGAGGACATGATCGGATATTAACCAAGCTTTCAGAGCCGCAAGG 303

QY 302 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTGTTTCATGCTATTTCACTCT 361
Db 304 CAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTGTTTCATGCTATTTCACTCT 363

QY 362 ATTAATCTTGCAACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGAAGCTTTC 421
Db 364 ATTAATCTTGCAACTGCACTCTTCTGCTCATGTTTCTTACGGCTTGAAGCTTTC 423

QY 422 GCTGCCATCCACACGCTGCTTTTGGCCGACCCGACCCGCGCTGACCTCCATCCCT 481
Db 424 GCTGCCATCCACACGCTGCTTTTGGCCGACCCGACCCGCGCTGACCTCCATCCCT 483

QY 482 CTGATCATGACGAGGTGTC 501
Db 484 TTGATTCACGACGAGTGTCC 503

RESULT 10
AX001030 1329 bp DNA linear PAT 10-MAR-2000
LOCUS AX001030
DEFINITION Sequence 12 from Patent WO9902666.
ACCESSION AX001030
VERSION AX001030.1 GI:7241266
KEYWORDS unclassified.
SOURCE unclassified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1329)
AUTHORS Ott, C. and Bedin, F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL Patent: WO 9902666-A 12 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
FEATURES
Location/Qualifiers
1..1329
/organism="unclassified"

BASE COUNT 346 a 352 c 288 g 340 t 3 others
ORIGIN
Query Match 85.0%; Score 426; DB 6; Length 1329;
Best Local Similarity 90.6%; Pred. No. 1e-126;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
/db_xref="taxon:32644"
QY 2 TCTGGGCCAAGCTCCCAAGACGACGATTTCTGTTGAGATGGGGAGCTGAGAGAC 61
DB 497 TGTACGCCAACCTCCCAAGACGATTTCTGTTGAGAGAGGAGCTGAGAGAC 556
QY 62 AGAAGTACCTGATTTCTAGAGCTGATTAAGATCCCTAAGCTTGGAGAGGTGAC 121
DB 557 AGAAGTACCTGATTTCTAGAGCTGATTAAGATCCCTAAGCTTGGAGAGGTGAC 616
QY 122 ACATTCACCTTTAAACAGGGGCTTGCACCTTATGTCACACTGACCAATCAGAGAGCTC 181
DB 617 GCATTCACCTTTAAACAGGGGCTTGCACCTTATGTCACACTGACCAATCAGAGAGCTC 676
QY 182 ACTAAATGCTAATTAGGCAAGAGGATTAAGAAATAGCAATGATCTATTCCTG 241
DB 677 ACTAAATGCTAATTAGGCAAGAGGATTAAGAAATAGCAATGATCTATTCCTG 736
QY 242 AGAGACACAGAGAGGACCAATGATCGGATATTAACCAAGTCTTTCAGCCGACG 301
DB 737 AGAGACACAGAGAGGACCAATGATCGGATATTAACCAAGTCTTTCAGCCGACG 796
QY 302 CAACCCCTTTGGGCTCCCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 361
DB 797 CAACCCCTTTGGGCTCCCTCTTTGATGGAGCTGTTTTCATGCTATTTCACTCT 856
QY 362 ATTAATCTGCACTGCACTCTTGGGCAATGTTCTTCAAGCTTGAAGCTTTC 421
DB 857 ATTAATCTGCACTGCACTCTTGGGCAATGTTCTTCAAGCTTGAAGCTTTC 916
QY 422 GCTCCGACACCACTGCTGTTTCCCGCACCGCAGCCGCTGATCCCATCTT 481
DB 917 GTTGGCCTACCACTGCTGTTTCCCGCACCGCAGCCGCTGATCCCATCTT 976
QY 482 CTGATCATGCAAGGTGTC 501
DB 977 TTGATCCAGCAGAGTGTTC 996
RESULT 11
AX000970 783 bp DNA linear PAT 10-MAR-2000
LOCUS AX000970 Sequence 15 from Patent WO9902696.
DEFINITION AX000970
ACCESSION AX000970
VERSION AX000970.1 GI:7241212
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 783)
TITLE Beseme, F. and Blond, J.
JOURNAL ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
AUTHORS DISEASES OR WITH PREGNANCY DISORDERS
PATENT: WO 9902696-A 15 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
SOURCE 1..783
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 173 a 213 c 166 g 180 t 51 others
ORIGIN
Query Match 85.0%; Score 425.6; DB 6; Length 783;
Best Local Similarity 94.0%; Pred. No. 1.3e-126;
Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;
QY 54 TGAGAGACAGACTAGCTGATTTCTAGCTGACTAAGATCCCTAAGCTTGAAGTGGGA 113

DB 1 TGAGAGACAGACTAGCTGATTTCTAGGCTGACTAAGATCCCTAAGCTTGAAGTGGGA 60
QY 114 AGTGACACATCCACTTTAAACACGGGGCTTGCACCTTATGCTCACACCTGACCAATCA 173
DB 61 AGTGACACATCCACTTTAAACACGGGGCTTGCACCTTATGCTCACACCTGACCAATCA 120
QY 174 GAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCACT 233
DB 121 GAGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCACT 180
QY 234 ATTGCTGAGAGCAACACAGAGGACAAATGATGGATTAATCCCAAGCTTGAAGC 293
DB 181 ATTGCTGAGAGCAACACAGAGGACAAATGATGGATTAATCCCAAGCTTGAAGC 240
QY 294 GGCAAGGCA-ACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTTTTCATGTA 352
DB 241 GGCAAGGCAAGMCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTTTTCATGTA 300
QY 353 TTTCATCTAATTAATCTTGCACCTGCTTGTGATGCTATTTTCAGGCTTGAAC 412
DB 301 TTTCATCTAATTAATCTTGCACCTGCTTGTGATGCTATTTTCAGGCTTGAAC 360
QY 413 TGAGCTTCCGCTCCGACCTCCACTGCTGTTTGGCCGACCGGACCGGCTGACT 472
DB 361 TGAGCTTCCGCTCCGACCTCCACTGCTGTTTGGCCGACCGGACCGGCTGACT 420
QY 473 CCCATCCCTGATCATGACGAGGTGTC 501
DB 421 CCCATCCCTGATCATGACGAGGTGTC 449
RESULT 12
HUA000660 251124 bp DNA linear PRI 15-JUN-2001
LOCUS HUA000660
DEFINITION Homo sapiens T-cell receptor alpha delta locus from bases 501613 to
752736 (section 3 of 5) of the Complete Nucleotide Sequence.
ACCESSION AE000660 AE000521 U85197
VERSION AE000660.1 GI:2358042
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 251124)
AUTHORS Koop, B.F., Rowen, L., Wang, K., Kuo, C.L., Seto, D., Lensz, J.A.,
Howard, S., Shan, W., Deshpande, P. and Hood, L.
TITLE The human T-cell receptor TCR α /TCR δ (C alpha/C delta) region:
organization, sequence, and evolution of 97.6 kb of DNA
JOURNAL Genomics 19 (3), 478-493 (1994)
MEDLINE 94245236
PUBMED 8188290
REMARK This citation covers from bases 966183-1064019
REFERENCE 2 (bases 1 to 251124)
AUTHORS Boysen, C., Simon, M.I. and Hood, L.
TITLE Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with
bacterial artificial chromosome clones
JOURNAL Genome Res. 7 (4), 330-338 (1997)
MEDLINE 97264339
PUBMED 9110172
REFERENCE 3 (bases 1 to 251124)
AUTHORS Boysen, C., Inyoub, L., Smith, T.M., Smit, A., Wang, K., Rowen, L. and
Hood, L.
TITLE T-cell Receptor Alpha Delta Locus Complete Nucleotide Sequence
JOURNAL Unpublished
REMARK This citation covers bases 1-983545 and bases 1064020-1071650
AUTHORS 4 (bases 1 to 251124)
Boysen, C., Lee, I., Smith, T.M., Smit, A., Kai, W., Lee, R. and Leroy, H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-1997) Department of Molecular Biotechnology,
University of Washington, Box 357730, Seattle, Washington 98195,
USA
REMARK Complete nucleotide sequence of the human T-cell receptor alpha

delta locus
On Jun 12, 2001 this sequence version replaced gi:2121231.
Sequencing method: high redundancy shotgun. Interspersed Repeats
were identified with RepeatMasker (available from
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>) Simple
sequence repeats were identified with sputnik (available from
<http://serac.mbc.washington.edu/~chrisa/software/sputnik.html>).
Location/Qualifiers

FEATURES
source
1. .251124
/organism="Homo sapiens"
/db_xref="taxon:9606"
157356. .23794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BAC 480"
/clone_lib="Mel Simon's BAC library"
/note="This span of sequence was submitted independently
as Genbank Accession Number U85197."
233572. .>251124
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BAC 378"
/clone_lib="Mel Simon's BAC library"
/note="This span of sequence was submitted independently
as Genbank Accession Number U85196."
complement(326. .479)
/rpt_family="MIR"
603. .983
/rpt_family="MLT1B"
1738. .2478
/gene="TCRAV17S1"
1738. .1757
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
join(<1957. .2008,2155. .>2439)
/gene="TCRAV17S1"
/product="TCRAV17S1"
/note="Old name TCRAV3S1"
join(1957. .2008,2155. .>2439)
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
/codon_start=1
/product="TCRAV17S1"
/protein_id="AAB69018.1"
/db_xref="GI:2358043"
/translation="METTLGSLVITLWQLARVNSQQEEDPOLSLIQEGENATMNC
SYKTSINMLQWYRQNSGRGLVHLILRSNERKHSGRLLVTLDTSKSSSLITASRAA
DTASVFCATD"
join(1957. .2008,2155. .>2439)
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="Old name TCRAV3S1"
2440. .2446
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="heptamer"
2447. .2469
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="spacer"
2470. .2478
/gene="TCRAV17S1"
/standard_name="TCRAV17S1"
/note="nonamer"
complement(4306. .4461)
/rpt_family="MIR2"
4975. .5275
/rpt_family="AluY"
complement(5366. .5574)
/rpt_family="MERS8A"
7244. .7983
gene

/gene="TCRAV18S1"
/pseudo
7244. .7263
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/pseudo
join(<7482. .7527,7656. .>7944)
/gene="TCRAV18S1"
/product="TCRAV18S1"
/pseudo
join(7482. .7527,7656. .>7944)
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/pseudo
7945. .7951
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="heptamer"
7952. .7974
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="spacer"
7975. .7983
/gene="TCRAV18S1"
/standard_name="TCRAV18S1"
/note="nonamer"
complement(8236. .8389)
/rpt_family="MERS9A"
complement(8556. .8740)
/rpt_family="L1MB7"
complement(9281. .10170)
/rpt_family="L1PA2"
complement(10023. .11230)
/rpt_family="L1"
11234. .11405
/rpt_family="L1"
11818. .12491
/gene="TCRAV19S1"
11818. .11837
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
join(<11900. .11948,12152. .>12452)
/gene="TCRAV19S1"
/product="TCRAV19S1"
/note="Old name TCRAV12S1"
join(11900. .11948,12152. .>12452)
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
/codon_start=1
/product="TCRAV19S1"
/protein_id="AAB69019.1"
/db_xref="GI:2358044"
/translation="MLTASLLRAVIASICVSSMAOKVTOAQTEISVVEREDVTLDCV
YETRDITYVLFWYKOPSPSELVFLIRNSPDEQNEIGRSYMNQKSTSSFNFTTAS
QVDSAVYFCALSEA"
join(11900. .11948,12152. .>12452)
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="Old name TCRAV12S1"
12453. .12459
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="heptamer"
12460. .12482
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="spacer"
12483. .12491
/gene="TCRAV19S1"
/standard_name="TCRAV19S1"
/note="nonamer"

FEATURES	* 160579	187321: contig of 26743 bp in length.
SOURCE	Location/Qualifiers	
	1. 187321	
	/organism="Papio cynocephalus anubis"	
	/sub_species="anubis"	
	/db_xref="taxon:9555"	
	/clone="RP41-107F15"	
	/clone_1ib="RP41"	
misc_feature	1. 25945	
	/note="assembly_fragment"	
	clone_end:r7	
	vector_side:left"	
misc_feature	1. 24997	
	/note="clone overlaps with GenBank Accession Number AC092516 clone RP41-76N20 (center project name ame)"	
misc_feature	26046. 59962	
	/note="assembly_fragment"	
misc_feature	60063. 77104	
	/note="assembly_fragment"	
misc_feature	77205. 160478	
	/note="assembly_fragment"	
misc_feature	103673. 187321	
	/note="clone overlaps with GenBank Accession Number AC096848 clone RP41-286K15 (center project name amg)"	
misc_feature	160579. 187321	
	/note="assembly_fragment"	
	clone_end:SP6	
	vector_side:right"	
BASE COUNT	56329 a 38211 c 38592 g 53789 t 400 others	
ORIGIN		
Query Match	81.8%; Score 409.6; DB 2; Length 187321;	
Best Local Similarity	92.3%; Pred. No. 4.1e-121;	
Matches 443; Conservative 0; Mismatches 34; Indels 3; Gaps 1;		
4	TCGGCAACCTCCCAACGACCTTAAGGTTTCTGTGAGATGGGGGACCTGAGACAG 63	
Db 54177	TCGGCAACCTCCCAATGACCTTAAGGTTTCTGTGAGAGGGGGACCTAAGACAG 54118	
64	GACTAGCTGATTTCTTAGCTGACTAAGAAATCCTTAAGCTAGCTGGGAAGTACAC 123	
Db 54117	GACTAGCTGATTTCTTAGGACAACTAAGAAATCCTTAAGCTAGCTGGGAAGTACAC 54058	
124	ATCCACCTTTAAACAGGGGCTTGAACCTAGCTCAACCTGACCAATCAGAGCTCAC 183	
Db 54057	ATCCACCTTTAAACAGGGGCTTGAACCTAGCTCAACCTGACCAATCAGAGCTCAC 53998	
184	TAAATGCTAATTAAGCAAGACAGAGGTAAAGAAATGCCAATCATTATTGCTGAG 243	
Db 53997	TAAATGCTAATTAAGCAAGACAGAGGTAAAGAAATGCCAATCATTATTGCTGAG 53938	
244	AGCAGACAGAGGAGCAATGATCGGGAATATTAACCCAAAGCTTTCGAGCGGCAACGCA 303	
Db 53937	AGCAGACAGGAGGAGCAAGAGATGGGGAATATAACCCAGGACTTCCAGCTGGCAACAGCA 53878	
304	ACCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTGTTTCATGCTAATTTCACTTA 363	
Db 53877	ACCCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTGTTTCATGCTAATTTCACTTA 53818	
364	TAAATCTTGCAATGACATCTTCTGTGTCATATGTTTCTTAAGCTGAGCTTTGGC 423	
Db 53817	TAAATCTTGCAATGACATCTTCTGTGTCATATGTTTCTTAAGCTGAGCTTTGGC 53758	
424	TCGCATCACCACAGCTGTGTCGCGCCACGACAGACCGGCGGTGACTCCCATCCCTCT 483	
Db 53757	TCGCATCACCACAGCTGTGTCGCGCCATCCGACAC--CCGCTGACTCCATCCCTCT 53701	
RESULT 14		
LOCUS	AC092843	83412 bp DNA linear PRI 21-FEB-2002
DEFINITION	Homoe sapiens BAC clone RP41-41904 from 2, complete sequence.	
ACCESSION	AC092843	

VERSION	AC092843.4	GI:18072221
KEYWORDS	HTG.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 83412)	
AUTHORS	Sulston,J.E. and Waterston,R.	
TITLE	Toward a complete human genome sequence	
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)	
MEDLINE	99063792	
PUBMED	9847074	
REFERENCE	2 (bases 1 to 83412)	
AUTHORS	Cedroni,M. and Haglund,K.	
TITLE	The sequence of Homo sapiens BAC clone RP11-41904	
JOURNAL	Unpublished (2001)	
REFERENCE	3 (bases 1 to 83412)	
AUTHORS	Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
REFERENCE	4 (bases 1 to 83412)	
AUTHORS	Waterston,R.H.	
TITLE	Direct Submission	
JOURNAL	Submitted (06-JUN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA	
REFERENCE	5 (bases 1 to 83412)	
AUTHORS	Waterston,R.	
TITLE	Direct Submission	
JOURNAL	Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Jan 6, 2002 this sequence version replaced gi:17921256.	
COMMENT		

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.-Y., Zhao, B., Frengen, E., Bateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

SOURCE INFORMATION: The RPEC-11 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Woon, P.-Y., Zhao, B., Frengen, E., Tatenoe, M., Carnease, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reesgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pRACE3.6

NEIGHBORING SEQUENCE INFORMATION:

SOURCE INFORMATION: The RPEC-11 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Woon, P.-Y., Zhao, B., Frengen, E., Tatenno, M., Carnease, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reesgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pRACE3.6

NEIGHBORING SEQUENCE INFORMATION:

|||||
Db 16276 AGCAGACGAGGAGGACAAAGATTGGATATATATCCAGGCATTCGAGCTGCACAGCA 16335
Oy 304 ACCCCCTTTGGGTCCCTCCCTTTGATGGAGAGCTGTGTTTCATGCTATTTTCACTTAT 363
Db 16336 ACCCCCTTTGGGTCCCTCCCTTTGATGGAGAG-CTGTTTTCACCTTATTTTCACTTAT 16393
Oy 364 TAAATCTTGCAACTGCACTCTTCTGGTCCATGTTTCTTAACGCTTGAGCTGAGCTTTCGC 423
Db 16394 TAAATCTTGCAACTGCACTCTTCTGGTGCATGTTTGTATACGCTTGAGCTGAACTTTCAC 16453
Oy 424 TCGCATTCACCACTGCTG-TTTGGCCGACCGCAGACCCGCGCTGACTCCATCCCTC 482
Db 16454 TCGCATTCACCACTGCTGTTTGGCGCGCTGCAGACCACTGCTGACTTCATTCCTTC 16513
Oy 483 TGGATCATGACGAGGTGCTCC 501
Db 16514 TGGATTCACGAGGTGCTCC 16532

RESULT 15
AL139038 140756 bp DNA linear PRI 15-JUL-2001
LOCUS AL139038
DEFINITION Human DNA sequence from clone RP11-456B18 on chromosome 13,
complete sequence.
ACCESSION AL139038
VERSION AL139038.18 GI:14800148
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140756)
Dunn,M.
Direct Submission
Submitted (14-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Jul 17, 2001 this sequence version replaced gi:13751261.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-456B18 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-456B18. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-456B18 is at 140756 in this
sequence. The true left end of clone RP11-78L16 is at 66441 in this
sequence. The true right end of clone RP11-108H9 is at 100 in this
sequence.
location/Qualifiers

FEATURES

source
1..140756
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-456B18"
/clone_id="RPC1-11.2"
1830..2326
/note="LOR1a repeat: matches 1..497 of consensus"
3932..3965
/note="17 copies 2 mer gt 94% conserved"
4632..5067
/note="LOR1b repeat: matches 1..461 of consensus"
5733..5893
/note="L1ME3 repeat: matches 6002..6146 of consensus"
6000..6083
/note="MER34 repeat: matches 454..539 of consensus"
6108..6171
/note="HERV23 repeat: matches 993..1056 of consensus"
7143..7840
/note="LTR8 repeat: matches 1..691 of consensus"
8059..8292
/note="LTR16a repeat: matches 203..442 of consensus"
8491..8686
/note="LTR29 repeat: matches 6..203 of consensus"
8725..8926
/note="MER4D repeat: matches 483..677 of consensus"
8992..9049
/note="MER4D repeat: matches 410..465 of consensus"
9062..9281
/note="110 copies 2 mer tt 59% conserved"
9398..9530
/note="MER41C repeat: matches 1..131 of consensus"
9623..9732
/note="LTR1 repeat: matches 675..785 of consensus"
9970..10333
/note="LTR2D repeat: matches 1..486 of consensus"
10400..10545
/note="LTR29 repeat: matches 2..151 of consensus"
10549..10935
/note="MER67D repeat: matches 12..391 of consensus"
11249..11332
/note="MIR repeat: matches 69..144 of consensus"
11923..12159
/note="MER4D repeat: matches 1..234 of consensus"
12164..12767
/note="MER4D repeat: matches 396..973 of consensus"
12852..13221
/note="THR1B repeat: matches 1..364 of consensus"
13222..14798
/note="THR1B-INTERNAL repeat: matches 1..1580 of
consensus"
14799..15173
/note="THR1B repeat: matches 1..364 of consensus"
16144..16346
/note="MER54B repeat: matches 585..789 of consensus"
16339..16503
/note="MER54B repeat: matches 112..278 of consensus"
16504..16571
/note="MER66A repeat: matches 1..69 of consensus"
16570..16629
/note="MER41B repeat: matches 283..341 of consensus"
16599..16902
/note="MER66A repeat: matches 163..478 of consensus"
16903..17014
/note="MER54B repeat: matches 3..115 of consensus"
17195..17491
/note="AluY repeat: matches 1..297 of consensus"
18120..19291
/note="L1p repeat: matches 1726..2897 of consensus"
19460..19584
/note="25 copies 5 mer attt 62% conserved"
19686..19745
/note="12 copies 5 mer tttaa 71% conserved"

```
repeat_region 20453. .20914
/notes="L1MD2 repeat: matches 5745. .6199 of consensus"
repeat_region 20913. .21222
/notes="L1MD2 repeat: matches 5277. .5594 of consensus"
repeat_region 21308. .21570
/notes="L1R9 repeat: matches 357. .625 of consensus"
repeat_region 21645. .21984
/notes="L1R19B repeat: matches 55. .393 of consensus"
repeat_region 22081. .22583
/notes="L1MD2 repeat: matches 4726. .5242 of consensus"
repeat_region 23498. .23784
/notes="L1ubd repeat: matches 1. .305 of consensus"
repeat_region 23873. .24169
/notes="L1usx repeat: matches 1. .295 of consensus"
repeat_region 25044. .25238
/notes="MIR repeat: matches 31. .238 of consensus"
repeat_region 25404. .25696
/notes="L1usx repeat: matches 20. .311 of consensus"
repeat_region 25832. .26218
/notes="MSTC repeat: matches 1. .405 of consensus"
repeat_region 26221. .27865
/notes="MSTC-internal repeat: matches 2. .1651 of consensus"
repeat_region 27907. .28659
/notes="L1P82 repeat: matches 5400. .6155 of consensus"
repeat_region 28710. .29043
/notes="MSTD repeat: matches 30. .394 of consensus"
repeat_region 29634. .29697
/notes="32 copies 2 mer t a 79% conserved"
repeat_region 29729. .29752
/notes="12 copies 2 mer t g 95% conserved"
repeat_region 30688. .30874
/notes="MER61A repeat: matches 160. .354 of consensus"
repeat_region 31648. .31785
/notes="L1MC4 repeat: matches 6409. .6540 of consensus"
repeat_region 31813. .32243
/notes="L1MC4 repeat: matches 6629. .7089 of consensus"
repeat_region 32449. .32821
/notes="THE1C repeat: matches 1. .371 of consensus"
repeat_region 32824. .34464
/notes="THE1C-internal repeat: matches 3. .1651 of consensus"
repeat_region 34465. .34825
/notes="THE1C repeat: matches 1. .360 of consensus"
repeat_region 34804. .35004
/notes="MER67C repeat: matches 202. .403 of consensus"
repeat_region 39792. .39868
/notes="L1P812 repeat: matches 6076. .6152 of consensus"
repeat_region 40953. .41258
/notes="L1y repeat: matches 1. .308 of consensus"
repeat_region 41279. .41308
/notes="15 copies 2 mer aa 86% conserved"
repeat_region 42744. .42933
/notes="L1MA8 repeat: matches 6064. .6256 of consensus"
repeat_region 42973. .43275
/notes="L1usx repeat: matches 1. .303 of consensus"
repeat_region 43991. .43991
/notes="match: STS: Em:HSPD07E5"
repeat_region 45196. .45584
/notes="MUT2B repeat: matches 2. .399 of consensus"
repeat_region 45594. .45767
/notes="87 copies 2 mer ta 77% conserved"
repeat_region 45800. .45851
/notes="MUT2B repeat: matches 394. .448 of consensus"
repeat_region 48118. .48487
/notes="THE1C repeat: matches 1. .370 of consensus"
repeat_region 48708. .48806
/notes="L1uio/FRAM repeat: matches 197. .297 of consensus"
repeat_region 49853. .50271
/notes="L1MEC repeat: matches 2274. .2369 of consensus"
repeat_region 50637. .50927
/notes="L1ME repeat: matches 974. .1253 of consensus"
repeat_region 52250. .52349
/notes="MIR repeat: matches 131. .228 of consensus"
```

```
repeat_region 52696. .52996
/notes="L1usx repeat: matches 1. .293 of consensus"
repeat_region 54927. .55010
Query Match 81.3% Score 407.2; DB 9; Length 14056;
Best local similarity 91.9%; Pred. No. 2.4e-120;
Matches 441; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
QY 4 TCGGCAACTGCCCAAGCAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGACAG 63
DB 69419 TTGGCAACTGCCCAAGCAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGACAG 69478
QY 64 GACTAGCTGATTTCTTCTGCTGCTAGTAAGATTCCTAGCTAGTGGAGGTACAC 123
DB 69479 GACTAGCTGATTTCTTCTGCTGCTAGTAAGATTCCTAGCTAGTGGAGGTACAC 69538
QY 124 ATCCACCTTTAAACGCGGGGCTGCACTTGTGCTACACCTGACCAATCAGAGACTGAC 183
DB 69539 ATCCACCTTTAAACGCGGGGCTGCACTTGTGCTACACCTGACCAATCAGAGACTGAC 69598
QY 184 TAAATGCTAATTAAGCAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 243
DB 69599 TAAATGCTAATTAAGCAAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 69658
QY 244 AGCAGCAGAGAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 303
DB 69659 AGCAGCAGAGAGAGAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 69718
QY 304 ACCGCTTTGGGTCCCTCCCTGTTGATAGGAGCTCTGTTGATGATTAATTAATTAATTAAT 363
DB 69719 ACCGCTTTGGGTCCCTCCCTGTTGATAGGAGCTCTGTTGATGATTAATTAATTAATTAAT 69778
QY 364 TAAATCTGCAACTGCACTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 422
DB 69779 TAAATCTGCAACTGCACTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 69838
QY 423 CTCGCCATCCACCACTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 482
DB 69839 CTCGCCATCCACCACTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 69898
```

Search completed: April 17, 2003, 06:48:14
Job time : 213.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 02:30:00 ; Search time 180.787 Seconds

(without alignments)
6240.778 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501
Sequence: 1 gtcgcgcacacccccca.....ctgcatcatgcaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10499	21	ABN97929
2	498	99.4	56093	24	ABL61744
3	461.4	92.1	7466	23	AA68626
4	450	89.8	711	21	ABN97947
5	426	85.0	1329	19	AAV43219
6	426	85.0	1329	20	AAK29704
7	425.6	85.0	783	20	AAK25669
8	396.2	79.1	711	21	ABN97946
9	396.2	79.1	1393	22	AAK31002

10	377.8	75.4	893	23	AA686964
11	377	75.2	2781	22	AA685630
12	375.4	74.9	2946	20	AAK77526
13	375.4	74.9	2946	21	AAK59468
14	374.4	74.7	2930	24	AAK24195
15	373.4	74.5	7582	20	AAK25665
16	373.4	74.5	7582	21	AAK59215
17	372.6	74.4	849	22	AAK51000
18	372.2	74.3	3372	20	AAK25663
19	372.2	74.3	3372	21	AAK59213
20	371.2	74.1	1136	20	AAK25660
21	371.2	74.1	1136	21	AAK59210
22	371.2	74.1	1136	23	AAK71727
23	371.2	74.1	5154	23	AAK67609
24	371.2	74.1	8279	23	AAK76474
25	371.2	74.1	8294	23	AAK84209
26	370.6	74.0	2782	20	AAK25661
27	370.6	74.0	2782	21	AAK59211
28	370.6	74.0	2782	22	AAK20069
29	350.8	70.0	2300	23	AAK92680
30	347.4	69.3	1165	23	AAK57271
31	346	69.1	2942	23	AAK77313
32	345.8	69.0	2629	22	AAK67491
33	341.8	68.2	635	19	AAK43215
34	341.8	68.2	635	20	AAK29702
35	341.8	68.2	2030	21	AAK63826
36	325.6	65.0	808	23	AAK88392
37	321.8	64.2	1478	23	AAK87568
38	320.8	64.0	1243	23	AAK84189
39	312.6	62.4	448	23	AAK87567
40	312.6	62.4	583	23	AAK88391
41	311.4	62.2	410	24	AAK94040
42	311.4	62.2	410	24	AAK64210
43	308.2	61.5	17758	22	AAK84138
44	303.4	60.6	439	22	AAK33993
45	303.4	60.6	439	22	AAK51671

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
AC ABN97929;
DT 01-AUG-2002 (first entry)
DE Human retroviral sequence HERV-7g.
KW Autoimmune disease; HERV-7g; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
OS Human retrovirus.
PN MO9967395-A1.
PD 29-DEC-1999.
PF 23-JUN-1999; 99WO-FR01513.
PR 23-JUN-1998; 98FR-0007920.
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI Alliel PM, Perin J, Rieger F;
DR WPI; 2000-160587/14.
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
XX used for diagnosis, treatment and prevention of autoimmune and
XX neurological diseases -
XX genomic sequence #

XX Claim 3; Fig 1; 225pp; French.
XX
PS The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q;
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;

Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 8,2e-161;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCTGGCCACCTCTCCCAACAGCACTTAGGTTTCTCTTTAGATGGGGAGTGAAGA 60
DB 9500 GTCTGGCCACCTCTCCCAACAGCACTTAGGTTTCTCTTTAGATGGGGAGTGAAGA 9559
QY 61 CAGGACTGCTGATTTCTTGAAGTGAAGATCCCTAAGCTTGAAGGAGTGAAGA 120
DB 9560 CAGGACTGCTGATTTCTTGAAGTGAAGATCCCTAAGCTTGAAGGAGTGAAGA 9619
QY 121 CACATCCACCTTTAAACAGCGGGCTTGCACCTAGCTCACCTGACCAATGAGAGCT 180
DB 9620 CACATCCACCTTTAAACAGCGGGCTTGCACCTAGCTCACCTGACCAATGAGAGCT 9679
QY 181 CACTAAATGCTTAATTAGGCAAGAGAGGTAAAGAAATGCCAATCATCTATTGCTT 240
DB 9680 CACTAAATGCTTAATTAGGCAAGAGAGGTAAAGAAATGCCAATCATCTATTGCTT 9739
QY 241 GAGACACAGCAGGAGGAGGCAATGATCGGATATTAACCAAGTCTTGAGCGGCAAGC 300
DB 9740 GAGACACAGCAGGAGGAGGCAATGATCGGATATTAACCAAGTCTTGAGCGGCAAGC 9799
QY 301 GCAACCCCTTTGGTGGTCCCTCTTTGATGAGGACCTCTGTTTCACTATTTCACCTC 360
DB 9800 GCAACCCCTTTGGTGGTCCCTCTTTGATGAGGACCTCTGTTTCACTATTTCACCTC 9859
QY 361 TATTAAATCTTGCAACTGCACTCTTGTGCAATGTTCTTAAGCGCTTGAAGCTTT 420
DB 9860 TATTAAATCTTGCAACTGCACTCTTGTGCAATGTTCTTAAGCGCTTGAAGCTTT 9919
QY 421 CGCTGGCCATCCACGACCTGCTTTGGCGGCAACCGCGGCGGCTGACCTCCATCCG 480
DB 9920 CGCTGGCCATCCACGACCTGCTTTGGCGGCAACCGCGGCGGCTGACCTCCATCCG 9979
QY 481 TCTGATCATGCGGGGTGTC 501
DB 9980 TCTGATCATGCGGGGTGTC 10000

KW gene; ds.
XX
OS Homo sapiens.
XX WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-235133P.
PR 18-SEP-2000; 2000US-235617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

(AVAIL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI, 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a
chemical agent to be tested for anti-neoplastic activity, and
determining a change in expression of a gene of a signature gene set
-
Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (i)
CC comprises a sequence (S) selected from 8447 sequences (given in AB161664
CC to AB170110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (i) has cytostatic

CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;

Query Match 99.4%; Score 498; DB 24; Length 56093;

Best Local Similarity 100.0%; Pred. No. 2,2e-159;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCCAAGCTCCCAACAGCAGCTTAGGTTTCTGTTGAGTGGGGAGCTGAGACAG 63
DB 37504 TCGGCCAAGCTCCCAACAGCAGCTTAGGTTTCTGTTGAGTGGGGAGCTGAGACAG 37563
QY 64 GACTAGCTGATTTCTGAGCTGATAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 123
DB 37564 GACTAGCTGATTTCTGAGCTGATAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 37623
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
DB 37624 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 37683
QY 184 TAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTGAG 243
DB 37684 TAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTGAG 37743
QY 244 AGCAGACAGAGAGGACATATGCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 303
DB 37744 AGCAGACAGAGAGGACATATGCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 37803
QY 304 ACCCCCTTTGGGTCCTCCCTCTTGTATGGAGACTCTGTTTCATGCTATTTCACCTAT 363
DB 37804 ACCCCCTTTGGGTCCTCCCTCTTGTATGGAGACTCTGTTTCATGCTATTTCACCTAT 37863
QY 364 TAAATCTTGCAACTGCACTCTTCTGTCATGTTTCTTAACGGCTTGAGTGAAGCTTTCGC 423
DB 37864 TAAATCTTGCAACTGCACTCTTCTGTCATGTTTCTTAACGGCTTGAGTGAAGCTTTCGC 37923
QY 424 TCGGCATCCACACTGCTGTTTGGCGCCACCGCAGACCCGCGCTGACTCCCATCCCTCT 483
DB 37924 TCGGCATCCACACTGCTGTTTGGCGCCACCGCAGACCCGCGCTGACTCCCATCCCTCT 37983
QY 484 GGATCATGACGGGTGTC 501
DB 37984 GGATCATGACGGGTGTC 38001

RESULT 3

AAS68626
ID AAS68626 standard; cDNA, 7466 BP.

XX AAS68626;

XX 13-FEB-2002 (first entry)

DB DNA encoding novel human diagnostic protein #4430.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG04439.

XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -

PS Claim 1; SEQ ID No 4430; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WRO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;

Query Match 92.1%; Score 461.4; DB 23; Length 7466;

Best Local Similarity 96.6%; Pred. No. 2.8e-147;
Matches 482; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

QY 4 TCGGCCAAGCTCCCAACAGCAGCTTAGGTTTCTGTTGAGTGGGGAGCTGAGACAG 63
DB 3668 TCGGCCAAGCTCCCAACAGCAGCTTAGGTTTCTGTTGAGTGGGGAGCTGAGACAG 37227
QY 64 GACTAGCTGATTTCTGAGCTGATAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 123
DB 3728 GACTAGCTGATTTCTGAGCTGATAGAAATCCCTAAGCCTAGCTGGGAAGTACAC 3787
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
DB 3788 ATCCACCTTTAAACAGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 3847
QY 184 TAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTGAG 243
DB 3848 TAAATGCTAATTAGGCAAGACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTGAG 3907
QY 244 AGCAGACAGAGAGGACATATGCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 303
DB 3908 AGCAGACAGAGAGGACATATGCGGATATTAACCAAGTCTTGAGACCGGCAACGGCA 3967
QY 304 -ACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTTTCATGCTATTTCACCTA 362
DB 3968 GCCCCTTTGGGTCCTCCCTCTTGTATGGAGCTCTGTTTCATGCTATTTCACCTA 4027
QY 363 TTAATCTTGCAACTGCACTCTTCTGTCATGTTTCTTAACGGCTTGAGTGAAGCTTTCG 422

```

Db 4028 TTTAACTTGGCACTGCTTCTGATCCATGTTCTTACGGCTCGAGCTGAGCTTTTG 4087
Qy 423 CTGGCATCCACACAGCTGTTTGGCGGACCGGAGACCCGGCGGTGACTCCATCCCTC 482
Db 4088 CTCACCGTCCACACAGCTGTTTGGCGGACCGGAGACCTGCGGTGACTCCATCCCTC 4147
Qy 483 TGGATCATGACAGGCTGTCC 501
Db 4148 TGGATCTCGACAGGTGTCC 4166

RESULT 4
ABN97947
ID ABN97947 standard; DNA; 711 BP.
AC ABN97947;
XX 01-AUG-2002 (first entry)
DE Human retroviral sequence R1F.
XX
XX Autocimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KM multiple sclerosis; ds.
XX
XX Human retrovirus.
OS
PN MO9967395-A1.
XX
XX 29-DEC-1999.
XX
XX 23-JUN-1999; 99WO-FR01513.
XX
XX 23-JUN-1998; 98FR-0007920.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Alliel PM, Perin J, Rieger F,
XX
XX WPI; 2000-160587/14.
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -
XX
XX Claim 3; Fig 3; 225pp; French.
XX
XX The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HIV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer.
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
XX Sequence 711 BP; 171 A; 209 C; 160 G; 171 T; 0 other;
XX
XX
XX Query Match 89.8%; Score 450; D3 21; Length 711;
XX Best local similarity 100.0%; Pred. No. 6.1e-144; Indels 0; Gaps 0;
XX Matches 450; Conservative 0; Mismatches 0;
XX
Qy 52 ACTGAGAGACAGAGTCTGATTTCTTAGGCTGACAAAGATCCCTAAGCTTAGCTGG 111
Db 1 ACTGAGAGACAGAGTCTGATTTCTTAGGCTGACAAAGATCCCTAAGCTTAGCTGG 60
Qy 112 GAAGTGACCACTCCATTTTAAACAGCGGGCTTGCACTTAGCTCACACCTGACCAAT 171
Db 61 GAAGTGACCACTCCATTTTAAACAGCGGGCTTGCACTTAGCTCACACCTGACCAAT 120

```

```

Qy 172 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCCATCAT 231
Db 121 CAGAGAGCTCACTAAATGCTAATTAGCAAGAAGAGAGGTAAAGAAATAGCCATCAT 180
Qy 232 CTATTGCTTGAAGACACAGCAGAGGAGCAATGATCGGAGATTAACCAAGCTTTGAG 291
Db 181 CTATTGCTTGAAGACACAGCAGAGGAGCAATGATCGGAGATTAACCAAGCTTTGAG 240
Qy 292 CCGGCAACGCAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTGTTTCATGCT 351
Db 241 CCGGCAACGCAACCCCTTTGGGTCCCTCCCTTTGATGGAGCTGTGTTTCATGCT 300
Qy 352 ATTTCACCTATTAATCTTGAACCTGACACTCTTCTGATCCATGTTTCTTAAGGCTTGA 411
Db 301 ATTTCACCTATTAATCTTGAACCTGACACTCTTCTGATCCATGTTTCTTAAGGCTTGA 360
Qy 412 CTGAGCTTTCGCTGGCCATCCACCACTGCTGTTGGCCGACCCGACAGCCGCGCTGAC 471
Db 361 CTGAGCTTTCGCTGGCCATCCACCACTGCTGTTGGCCGACCCGACAGCCGCGCTGAC 420
Qy 472 TCCCATCCCTCGATCATGAGGCTGTCC 501
Db 421 TCCCATCCCTCGATCATGAGGCTGTCC 450

RESULT 5
AAV43219
ID AAV43219 standard; CDNA; 1329 BP.
AC AAV43219;
XX
XX 29-DEC-1998 (first entry)
XX
XX Multiple sclerosis associated retrovirus fragment 6.
XX
XX Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
KM gag gene; env gene; rheumatoid arthritis-associated virus; ss.
XX
XX Multiple sclerosis associated retrovirus.
OS
XX
XX Key Location/Qualifiers
XX CDS 2..490
XX
XX /*tag= a
XX /product= "Encodes protein AAV71069"
XX /transl_except= (pos:77-79, appears to code for a
XX stop codon)
XX /transl_except= (pos:125-127, appears to code for a
XX stop codon)
XX /transl_except= (pos:137-139, appears to code for a
XX stop codon)
XX
XX WO9823755-A1.
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-IB01482.
XX
XX 26-NOV-1996; 96US-0756429.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F,
XX Mandrand B, Paranhos-Baccala G, Perron H;
XX
XX WPI; 1998-322732/28.
XX
XX P-PSDB; AAV71069.
XX
XX New nucleic acid from retroviruses - useful for diagnosis,
XX prevention and treatment of, e.g. multiple sclerosis
XX
XX Disclosure; Pages 187-188; 286pp; English.
XX
XX

```

CC The present sequence represents a multiple sclerosis (MS) associated
CC retrovirus (MSRV) genomic fragment used in the method of the
CC invention. The invention provides complete or partial genomic
CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
CC polypeptides encoded by these genes. The invention also provides
CC antibodies raised against the polypeptides. The genomic sequences,
CC polypeptides and antibodies are also claimed useful for diagnosing
CC infection by MS and rheumatoid arthritis-associated viruses, and also
CC for prevention and treatment of infection with these viruses.

XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 19; Length 1329;

Best Local Similarity 90.6%; Pred. No. 1.6e-135; Mismatches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCCACTCCCAAGCACTTGTGTTTCTGTTGAGTGGGAGCTGAGAC 61
DB TGTGAGCCAACTCCCAAGCACTTGTGTTTCTGTTGAGTGGGAGCTGAGAC 556
QY 62 AGAAGTACTGATTTCTAGGCTGATAGAATCCCTAAGCTTGGAGAGTGAC 121
DB 557 AGGACTACTGATTTCTAGGCTGATAGAATCCCTAAGCTTGGAGAGTGAC 616
QY 122 ACATCCACTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGCTC 181
DB 617 GCATCCATCTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGCTC 676
QY 182 ACTAAATGCTTAATAGCAAGAGGTAAGAAATAGCAATCATCTATTGCTG 241
DB 677 ACTAAATGCTTAATAGCAAGAGGTAAGAAATAGCAATCATCTATTGCTG 736
QY 242 AGAGCAGCAGAGAGGGAATGATCGGATATAAACCCTTGAACCGCGAGCG 301
DB 737 AGAGCAGCAGAGAGGGAATGATCGGATATAAACCCTTGAACCGCGAGCG 796
QY 302 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTGTTTCACTATTGACTCT 361
DB 797 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTGTTTCACTATTGACTCT 856
QY 362 ATTAATCTTGAATGCACTTTCTGTCATGTTTCTTAAGGCTTGAAGCTTTC 421
DB 857 ATTAATCTTGAATGCACTTTCTGTCATGTTTCTTAAGGCTTGAAGCTTTC 916
QY 422 GCTGCGCATCCACCACTGCTGTTTGGCGCACCGGACCGCGTGAATCCCT 481
DB 917 GTTCCGATCCACCACTGCTGTTTGGCGCACCGGACCGCGTGAATCCCT 976
QY 482 CTGATCATGACGAGGTGTC 501
DB 977 TTGATCCAGAGAGTGTCC 996

RESULT 6
AA29704
ID AA29704 standard; DNA; 1329 BP.

XX AAX29704;

AC AAX29704;

XX 08-JUN-1999 (first entry)

DE Clone 5M6 from MSRV-1.

XX Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
KM rheumatoid polyarthritis; ss.

XX Multiple sclerosis related virus type 1.

XX FR2765588-A1.

XX 08-JAN-1999.

XX 07-JUL-1997; 97FR-000816.

XX 07-JUL-1997; 97FR-000816.

XX (INMR) BIO MERIEUX.

XX WPI: 1999-098275/09.

XX P-PSDB; AAM99554.

XX Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis

XX Claim 1; Page 39-40; 83pp; French.

CC This sequence represents clone 5M6 from a novel multiple sclerosis
CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with
CC rheumatoid polyarthritis.

XX Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 85.0%; Score 426; DB 20; Length 1329;

Best Local Similarity 90.6%; Pred. No. 1.6e-135; Mismatches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 2 TCTGGCCCACTCCCAAGCACTTGTGTTTCTGTTGAGTGGGAGCTGAGAC 61
DB 497 TGTGAGCCAACTCCCAAGCACTTGTGTTTCTGTTGAGTGGGAGCTGAGAC 556
QY 62 AGAAGTACTGATTTCTAGGCTGATAGAATCCCTAAGCTTGGAGAGTGAC 121
DB 557 AGGACTACTGATTTCTAGGCTGATAGAATCCCTAAGCTTGGAGAGTGAC 616
QY 122 ACATCCACTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGCTC 181
DB 617 GCATCCATCTTTAAACAGCGGGCTTGAACCTTGAACCTGACCAATCAGAGCTC 676
QY 182 ACTAAATGCTTAATAGCAAGAGGTAAGAAATAGCAATCATCTATTGCTG 241
DB 677 ACTAAATGCTTAATAGCAAGAGGTAAGAAATAGCAATCATCTATTGCTG 736
QY 242 AGAGCAGCAGAGAGGGAATGATCGGATATAAACCCTTGAACCGCGAGCG 301
DB 737 AGAGCAGCAGAGAGGGAATGATCGGATATAAACCCTTGAACCGCGAGCG 796
QY 302 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTGTTTCACTATTGACTCT 361
DB 797 CAACCCCTTTGGGTCCTCCCTCTTGTATGGAGCTGTTTCACTATTGACTCT 856
QY 362 ATTAATCTTGAATGCACTTTCTGTCATGTTTCTTAAGGCTTGAAGCTTTC 421
DB 857 ATTAATCTTGAATGCACTTTCTGTCATGTTTCTTAAGGCTTGAAGCTTTC 916
QY 422 GCTGCGCATCCACCACTGCTGTTTGGCGCACCGGACCGCGTGAATCCCT 481
DB 917 GTTCCGATCCACCACTGCTGTTTGGCGCACCGGACCGCGTGAATCCCT 976
QY 482 CTGATCATGACGAGGTGTC 501
DB 977 TTGATCCAGAGAGTGTCC 996

RESULT 7

AA25669
ID AA25669 standard; cDNA to mRNA; 783 BP.

XX AAX25669;

XX 21-MAY-1999 (first entry)

XX Human endogenous retrovirus W long terminal repeat region.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 OS Human endogenous retrovirus.

XX MO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98MO-FR01442.

XX 07-JUL-1997; 97FR-0008015.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 83; 106pp; French.

XX This sequence represents the long terminal repeat (LTR) region of the
 CC human endogenous retrovirus (HERV) W genome. The nucleic acids, their
 CC fragments or peptides encoded by them are markers of autoimmune disease
 CC (e.g. multiple sclerosis, Rheumatoid polyarthritis, disseminated lupus
 CC erythematosus, insulin-dependent diabetes and related pathologies) and
 CC of abnormal or unsuccessful pregnancy and can be used as chromosomal
 CC makers for susceptibility to these conditions, or proximity markers
 CC of genes associated with this susceptibility.

XX Sequence 783 BP; 173 A; 213 C; 166 G; 180 T; 51 other;

XX Query Match 85.0%; Score 425.6; DB 20; Length 783;

XX Best Local Similarity 94.0%; Pred. No. 1.6e-135; Mismatches 1; Indels 1; Gaps 1;

XX Matches 422; Conservative 25; Mismatches 1; Indels 1; Gaps 1;

XX 54 TGAGAGACAGAGCTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTTACCTGGGA 113

XX 1 TGAGAGACAGAGCTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTTACCTGGGA 60

XX 114 AGGTACCAACATCCACTTTAAACAGGGGCTTGCAACTAGCTCAGCTGACCAATCA 173

XX 61 AGGTACCAACATCCACTTTAAACAGGGGCTTGCAACTAGCTCAGCTGACCAATCA 120

XX 174 GAGAGCTCACTAAATGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATCT 233

XX 121 GAGAGCTCACTAAATGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATCT 180

XX 234 ATTGCCTGAGACACAGCAGAGGACAATGATCGGATATTAACCAAGCTTGGAGCC 293

XX 181 ATTGCCTGAGACACAGCAGAGGACAATGATCGGATATTAACCAAGCTTGGAGCC 240

XX 294 GGCAACGGCA-AACCCCTTTGGGTCCCTCCCTTTGATGGAGGCTTTTCAATGCTA 352

XX 241 GGCAACGGCA-AACCCCTTTGGGTCCCTCCCTTTGATGGAGGCTTTTCAATGCTA 300

XX 353 TTTCACCTATTAATCTTGAACACTGCTCTGATTCATAGTTTCTTAACGGCTTGAAC 412

XX 301 TTTCACCTATTAATCTTGAACACTGCTCTGATTCATAGTTTCTTAACGGCTTGAAC 360

XX 413 TGAGCTTTGCTGCGCATCACCACCTGCTGTTTGGCGACACGACAGACCCGCGCTGACT 472

XX 361 TGAGCTTTGCTGCGCATCACCACCTGCTGTTTGGCGACACGACAGACCCGCGCTGACT 420

XX 473 CCCATCCCTTGATCATGACGAGGTGCC 501

XX 421 CCCATCCCTTGATCATGACGAGGTGCC 449

RESULT 8

ID ABR97946 standard; DNA; 711 BP.

XX ABR97946;

XX 01-AUG-2002 (first entry)

XX Human retroviral sequence R1.

XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;

XX multiple sclerosis; ds.

XX Human retrovirus.

XX MO9967395-A1.

XX 29-DEC-1999.

XX 23-JUN-1999; 99MO-FR01513.

XX 23-JUN-1998; 98FR-0007920.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX Alliel PM, Perin J, Rieger F;

XX WPI; 2000-160587/14.

XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
 PT used for diagnosis, treatment and prevention of autoimmune and
 PT neurological diseases -

XX Claim 3; Fig 3; 225pp; French.

XX The present invention relates to new nucleic acid sequences of human
 CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
 CC Regulatory elements associated with HERV-7q may alter expression of other
 CC genes (even remote genes) on the same chromosome, inducing immunological
 CC and/or neurological changes (which may be pathological or protective/
 CC curative). HERV-7q peptides can be used to improve efficiency of the
 CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
 CC sequences can be used in immunogenic or vaccinating compositions, for
 CC protection against autoimmune diseases, particularly multiple sclerosis.
 CC The peptides may also be used (by sequence comparison) to detect/identify
 CC endogenous retroviruses that are abnormally expressed in cancer,
 CC neuropathologies or other autoimmune diseases. The present sequence was
 CC used to illustrate the invention.

XX Sequence 711 BP; 170 A; 204 C; 162 G; 175 T; 0 other;

XX Query Match 79.1%; Score 396.2; DB 21; Length 711;

XX Best Local Similarity 95.3%; Pred. No. 2e-125; Mismatches 18; Indels 3; Gaps 2;

XX Matches 430; Conservative 0; Mismatches 18; Indels 3; Gaps 2;

XX 52 ACTGAGACAGAGCTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTTACCTGG 111

XX 1 ACTGAGACAGAGCTAGCTGATTTCTTAGCTGACTAAGAAATCCCTAAGCTTACCTGG 60

XX 112 GAAGTGACCAATCACCCTTTAAACACGGGGCTTGCACTTACCTCAGCTGACCAAT 171

XX 61 GAAGTGACCAATCACCCTTTAAACACGGGGCTTGCACTTACCTCAGCTGACCAAT 120

XX 172 CAGAGAGCTCACTAAATGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATCT 231

XX 121 CAGAGAGCTCACTAAATGCTAATTAGCAAGACAGAGGTAAAGAAATAGCCATCATCT 180

XX 232 CTATTGCTGAGACACAGCAGAGGACAATGATCGGATATTAACCAAGCTTGGAG 291

XX 181 CTATTGCTGAGACACAGCAGAGGACAATGATCGGATATTAACCAAGCTTGGAG 240

XX 292 CCGGCAACGGCA-AACCCCTTTGGGTCCCTCCCTTTGATGGAGGCTTCTGTTTCAATGC 350

Db 547 TCACCTATTAATCTTCACATGCACTCTCTGCTGTGTGTTGTAAGGTTAGCTG 606
Qy 415 AGCTTTGCTGCTCCATCCACGACCTGTTGGCCGCAACCGAGACCGCGCTGACCC 474
Db 607 AGCTTTGCTGCTCCATCCACGACCTGTTGGCCGCAACCGAGACCTGCTGACCTTC 666
Qy 475 CATCCCTGATCATGACGAGGTGTCC 501
Db 667 CATCCCTCGGATCTGGCAGGCTGTTC 693

RESULT 10
AAS65964
ID AAS65964 standard; cDNA; 893 BP.
XX
AC AAS65964;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #1768.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABE01777.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 1768; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.

Sequence 893 BP; 223 A; 264 C; 202 G; 204 T; 0 other;

Query Match 75.4%; Score 377.8; DB 23; Length 893;
Best Local Similarity 88.6%; Pred. No. 4.7e-119;
Matches 451; Conservative 0; Mismatches 37; Indels 21; Gaps 3;
Qy 4 TCGGCCAACCTCCCAACAGCACTTAGTTTCTCTGTGAGATGGGGGACGAGAGAG 63
Db 188 TCGGCCAACCTCCCAACAGCACTTAGTTTCTCTGTGAGATGGGGGACGAGAGAG 247
Qy 64 GACTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCCTAGCTGGAGAGTACAC 123
Db 248 GACTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCCTAGCTGGAGAGTACAC 307
Qy 124 ATCCACCTTTAAACACGGGGCTTCAACTTAGTCACTGACCAATC-----AG 174
Db 308 ATCCACCTTTAAACACGGGTGCTTCAACTGACCTACACCCGACCATCAGTAAGAAAG 367
Qy 175 AGAGCTCACTAAATGCTAATTTAGCAAAAGCAGAGGTAAAGAAATGCAATCATCTA 234
Db 368 AGAGCCCGCTAAATGCTAATTTAGCAAAAGCAGAGGTAAAGAAATGCAATCATCTA 427
Qy 235 TTGCCTGAGACAGACAGAGAGCAATGATCGGATATTAACCAAGTCTTGAGCCG 294
Db 428 TTGCCTGAGACAGACAGAGAGCAATGATCAAGATATTAACCAAGTCTTGAGCCG 487
Qy 295 GCAAGCGCAACCCCTTTGGGTCCCTCTTGTATGAGAGCTGTGTTTCATGCTATT 354
Db 488 GCAAGCACTACCTCTTTGGGTCCCTCTTGTATGAGAGCTGTG-----TT 537
Qy 355 TCACCTATTAATCTTGCACATG--CACTCTTGTGCTCAATGTTCTTAGGCTTAGC 412
Db 538 TCACCTATTAATCTTGCACATGCACTCTTGTGCTGTGTGTTGTTAGGCTCGAGC 597
Qy 413 TGAGCTTTCGCTCGCATCCACCACTGCTGTTGGCCGACCGAGACCCCGCTGACT 472
Db 598 TGAGCTTTCGCTCAACCGTCCACCACTGCTGTTGGCCACCGTGGACAGACCTGACT 657
Qy 473 CCCATCCCTGTGATCATGACGGGTGTCC 501
Db 658 TCCACCCCTCCAGATATGACGAGGGTGTCC 686

RESULT 11
AAS55630
ID AAS55630 standard; DNA; 2781 BP.
XX
AC AAS55630;
XX
DT 29-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
XX Envelope protein; HERV; syncytia formation; placental development;
XX syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 762..2378
FT /tag= a
FT /product= "envelope protein"
PN WO200116171-A1.
XX
PD 08-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-FR02429.
XX
PR 01-SEP-1999; 99FR-0011141.
PR 15-SEP-1999; 99FR-0011793.
XX
XX (INNR) BIO MERIEUX.
PA (INNR) INST NAT SANTE & RECH MEDICALE.

XX Mallet F, Cosset F, Blond J, Lavillette D, Boulton O, Ruggieri A;
PI WPI; 2001-226676/23.
XX P-PSDB; AAB67652.
DR
XX
PT Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia -
XX
PS Disclosure; Page 44-45; 57pp; French.
XX
CC The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV), in
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.
XX
SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;

Query Match 75.2%; Score 377; DB 22; Length 2781;
Best Local Similarity 100.0%; Pred. No. 1,6e-118;
Matches 377; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAAGACGACCTTAGGTTTCTGTGAGATGGGGGACTGAGAGACAG 63
Db 2387 TCGGCAACCTCCCAAGACGACCTTAGGTTTCTGTGAGATGGGGGACTGAGAGACAG 2446

QY 64 GACTAGCTGATTTCTCTAGGCTGATAGAAATCCCTAAGCCTAGCTGGAGAGTACAC 123
Db 2447 GACTAGCTGATTTCTCTAGGCTGATAGAAATCCCTAAGCCTAGCTGGAGAGTACAC 2506

QY 124 ATCCACCTTTTAAACAGCGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAG 183
Db 2507 ATCCACCTTTTAAACAGCGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAG 2566

QY 184 TAAATGCTAATTAGGCAAGACAGAGGTTAAAGAAATAGCCATCATCTATTGCTTGG 243
Db 2567 TAAATGCTAATTAGGCAAGACAGAGGTTAAAGAAATAGCCATCATCTATTGCTTGG 2626

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCGGCAACGGCA 303
Db 2627 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCGGCAACGGCA 2686

QY 304 ACCGCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 363
Db 2687 ACCGCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 2746

QY 364 TAAATCTTGCACTGCA 380
Db 2747 TAAATCTTGCACTGCA 2763

RESULT 12
AA77526
ID AA77526 standard; cDNA; 2946 BP.
XX
AC AA77526;
XX
DT 10-AUG-1999 (first entry)
DE Human secreted protein AJ172_2 cDNA.
XX
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;

KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX
OS Homo sapiens.
XX W09926972-A1.
XX
PD 03-JUN-1999.
XX
PF 17-NOV-1998; 98WO-US24614.
XX
PR 20-OCT-1998; 98US-0175928.
PR 21-NOV-1997; 97US-0976110.
PR 18-MAY-1998; 98US-0080478.
XX
PA (GENY) GENETICS INST INC.
PI Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;
PI McCoy JM, Werberg D, Treacy M;
XX
DR WPI; 1999-357813/30.
DR P-PSDB; AAY08622.
XX
FT New polynucleotides encoding secreted proteins
XX
PS Claim 13a; Page 100-101; 142pp; English.
XX
CC This invention describes novel human secreted proteins encoded by
CC polynucleotides isolated from human adult testes, adult brain, adult
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC libraries. The products of the invention are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals, although no
CC supporting data is given. Suggested activities include nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating (e.g. as vaccines) or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. The polynucleotides are also stated to be useful
CC for gene therapy.
XX
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

Query Match 74.9%; Score 375.4; DB 20; Length 2946;
Best Local Similarity 99.7%; Pred. No. 6e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAAGACGACCTTAGGTTTCTGTGAGATGGGGGACTGAGAGACAG 63
Db 2553 TCGGCAACCTCCCAAGACGACCTTAGGTTTCTGTGAGATGGGGGACTGAGAGACAG 2612

QY 64 GACTAGCTGATTTCTCTAGGCTGATAGAAATCCCTAAGCCTAGCTGGAGAGTACAC 123
Db 2613 GACTAGCTGATTTCTCTAGGCTGATAGAAATCCCTAAGCCTAGCTGGAGAGTACAC 2672

QY 124 ATCCACCTTTTAAACAGCGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAG 183
Db 2673 ATCCACCTTTTAAACAGCGGGGCTTGGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAG 2732

QY 184 TAAATGCTAATTAGGCAAGACAGAGGTTAAAGAAATAGCCATCATCTATTGCTTGG 243
Db 2733 TAAATGCTAATTAGGCAAGACAGAGGTTAAAGAAATAGCCATCATCTATTGCTTGG 2792

QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCGGCAACGGCA 303
Db 2793 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCCAAAGTCTTGGAGCGGCAACGGCA 2852

QY 304 ACCGCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 363
Db 2853 ACCGCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTCACTCTAT 2912

Qy 364 TAAATCTTGCAACTGCA 380
 Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 13

AAZ59468
 ID AAZ59468 standard; cDNA; 2946 BP.

AC AAZ59468;

DT 11-APR-2000 (first entry)

DE Human secreted protein AJ172_2 polynucleotide sequence.

XX Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
 KM placental pathology; metastasis inhibition; nutritional activity;
 KM immune stimulator; haematopoiesis regulator; tissue growth;
 KM tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
 gene therapy; ss.

XX Homo sapiens.

PN WO9960020-A1.

PD 25-NOV-1999.

PF 17-MAY-1999; 99WO-US10915.

PR 18-MAY-1998; 98US-0080478.

PR 20-OCT-1998; 98US-0175928.

PA (GENY) GENETICS INST INC.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Mi S, Treacy M;

DR WPI; 2000-116311/10.

DR P-PSDB; AAY67313.

PT New polynucleotides encoding secreted cDNA libraries, used to develop

PT products for the diagnosis and treatment of neoplastic disease -

PS Claim 14; Page 107-108; 149pp; English.

XX This is the human secreted protein AJ172_2 nucleotide sequence, obtained
 CC from a human adult testes cDNA library. The invention relates to secreted
 CC human and murine proteins. The polynucleotides and proteins are predicted
 CC to have biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals. Detection of the levels of the proteins can be used for the
 CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
 CC which modulate the expression or function of the proteins may be used for
 CC treating a neoplastic disease and inhibiting metastasis. Other suggested
 CC activities include nutritional activity (e.g. in feeds), cytokine and
 CC cell proliferation/differentiation activity, immune stimulating
 CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
 CC activity, tissue growth activity, activin/inhibin activity,
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
 CC invasion suppressor activity, and tumour inhibition activity. The
 CC polynucleotide sequences are also stated to be useful for gene therapy.

XX Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

Query Match 74.9%; Score 375.4; DB 21; Length 2946;

Best Local Similarity 99.7%; Pred. No. 66-118;

Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TCGGCCAACCTCCCAACGACCTAGTTCCTGCTGAGATGGGGGACTGAGAGACAG 63

Db 2553 TCGGCCAACCTCCCAACGACCTAGTTCCTGCTGAGATGGGGGACTGAGAGACAG 2612

Qy 64 GACTAGCTGCATTTCTGAGGCTGACTAAGATTCCTTAAGCTTACTGGGAAGTGACCC 123
 Db 2613 GACTAGCTGCATTTCTGAGGCTGACTAAGATTCCTTAAGCTTACTGGGAAGTGACCC 2672
 Qy 124 ATCCACCTTTAAACGGGGGCTTCAACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 183
 Db 2673 ATCCACCTTTAAACGGGGGCTTCAACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 2732
 Qy 184 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCAAATCATATTGCTTAG 243
 Db 2733 TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCAAATCATATTGCTTAG 2792
 Qy 244 AGCAGCAGAGAGGACATATATCGGATATAAACCAAGTCTTGAGCCGCGAACGCA 303
 Db 2793 AGCAGCAGAGAGGACATATATCGGATATAAACCAAGTCTTGAGCCGCGAACGCA 2852
 Qy 304 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTGTTTTCATGCTATTTCCTAT 363
 Db 2853 ACCCCCTTGGGTCCTCCCTTGTATGGAGCTGTTTTCATGCTATTTCCTAT 2912
 Qy 364 TAAATCTTGCAACTGCA 380
 Db 2913 TAAATCTTGCAACTGCA 2929

RESULT 14

AAD24195
 ID AAD24195 standard; cDNA; 2930 BP.

AC AAD24195;

DT 07-MAY-2002 (first entry)

DE Human syncytin cDNA.

XX Human; syncytin; pre-eclampsia; gestational trophoblast disorder;

KM choriocarcinoma; hydatiform mole; placental site tumour; abortion;

KM envelope gene; human endogenous defective retrovirus; HBRV-W; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS /30..2546

FT /tag= a

FT /product= "Syncytin"

XX MO200204678-A2.

XX 17-JAN-2002.

XX 09-JUL-2001; 2001WO-US21719.

XX 07-JUL-2000; 2000US-216657P.

XX (GENY) GENETICS INST INC.

XX Keith JC, McCoy JM, Mi S;

XX WPI; 2002-171727/22.

XX P-PSDB; AAE14540.

XX Identifying a compound for treating a subject with or at risk of

XX developing pre-eclampsia, comprises determining whether the expression

XX of activity of syncytin in the cell is modulated in the presence of a

XX test compound -

XX Disclosure; Page 39-42; 43pp; English.

XX The invention relates to identifying compounds which are modulators

XX of syncytin expression. The syncytin modulators are useful in diagnosis

XX and treatment of pre-eclampsia and gestational trophoblast disorders (e.g.

XX choriocarcinoma, hydatiform mole, placental site tumour and missed/

XX incomplete abortion). Syncytin is a human gene derived from the

CC envelope gene of human endogenous defective retrovirus, HERV-W. The
CC present invention is based partly on the discovery that syncytin
CC expression is dramatically reduced in preclampsia, and is also
CC mis-localised to the apical syncytiotrophoblast membrane. The present
CC sequence is human syncytin cDNA.

XX Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;

Query Match 74.7%; Score 374.4; DB 24; Length 2930;
Best Local Similarity 99.7%; Pred. No. 1.3e-117;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 63
Db TCGGCCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 2614
QY 64 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTGAGGAGGTGACAC 123
Db 2615 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTGAGGAGGTGACAC 2674
QY 124 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
Db 2675 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 2734
QY 184 TAAATGCTAATTAGGCAAGAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 243
Db 2735 TAAATGCTAATTAGGCAAGAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 2794
QY 244 AGCAGACAGAGGAGGACATGATCGGGATATTAACCAAGTCTTGAGCCGGCAACGGCA 303
Db 2795 AGCAGACAGAGGAGGACATGATCGGGATATTAACCAAGTCTTGAGCCGGCAACGGCA 2854
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGATGAGGAGCTGTGTTTCATGCTATTACCTCTAT 363
Db 2855 ACCCCCTTTGGGTCCCTCCCTTTGATGAGGAGCTGTGTTTCATGCTATTACCTCTAT 2914
QY 364 TAAATCTTGAACCTGC 379
Db 2915 TAAATCTTGAACCTGC 2930

RESULT 15
AAK25665
ID AAK25665 standard; cDNA to mRNA; 7582 BP.

XX AAK25665;

XX 21-MAY-1999 (first entry)

XX Complete human endogenous retrovirus W genome.

KW Clone; human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; 88.

XX Human endogenous retrovirus.

XX MO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INMR) BIO MERIEUX.

XX Beseme F, Blond JL, Boucon O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

PT New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and

PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
XX Claim 1; Page 71-74; 106pp; French.

CC This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin-dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.

XX Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 74.5%; Score 373.4; DB 20; Length 7582;
Best Local Similarity 97.6%; Pred. No. 4.9e-117;
Matches 368; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCGGCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 63
Db 7206 TCGGCCAACCTCCCAACAGCACTTAGGTTTCTGTGAGATGGGGAGCTGAGAGACAG 7265
QY 64 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTGAGGAGGTGACAC 123
Db 7266 GACTAGCTGATTTCTTGAGCTGACTAAGATCCCTAAGCCCTGAGGAGGTGACAC 7325
QY 124 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 183
Db 7326 ATCCACCTTTAAACACGGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGACTCAG 7385
QY 184 TAAATGCTAATTAGGCAAGAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 243
Db 7386 TAAATGCTAATTAGGCAAGAAGACAGAGGTAAAGAAATAGCCCAATCATCTATTGCTGAG 7445
QY 244 AGCAGACAGAGGAGGACATGATCGGGATATTAACCAAGTCTTGAGCCGGCAACGGCA 303
Db 7446 AGCAGACAGAGGAGGACATGATCGGGATATTAACCAAGTCTTGAGCCGGCAACGGCA 7505
QY 304 ACCCCCTTTGGGTCCCTCCCTTTGATGAGGAGCTGTGTTTCATGCTATTACCTCTAT 363
Db 7506 ACCCCCTTTGGGTCCCTCCCTTTGATGAGGAGCTGTGTTTCATGCTATTACCTCTAT 7565
QY 364 TAAATCTTGAACCTGC 380
Db 7566 TAAATCTTGAACCTGC 7582

Search completed: April 17, 2003, 05:07:30
Job time : 206.787 secs

Qy 304 ACCCCCTTGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTAT 363
Db 2853 ACCCCCTTGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCACCTAT 2912
Qy 364 TAAATCTTGCACTGCA 380
Db 2913 TAAATCTTGCACTGCA 2929

RESULT 2

US-08-686-878A-50
; Sequence 50, Application US/08686878A
; Patent No. 5708157
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Werberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/686,878A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-686-878A-50

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

Qy 119 ACCACATCCACTTTAAACAGCGGGCTTGCACTTACCTCACCTGACCAATCAGAGAG 178
Db 1 RCACATCCACCTTTAAACAGCGGGCTTGCAANAAGATNACCTTGACCAATCAGAGAG 60
Qy 179 CTCACATAATGCTAATTAGGCAAGAGAGGTAAAGAAATAGCCATCTATTTCG 238
Db 61 NTCANTAAATATATATTNGGCAAAAAGAGAGTAAAGAAATAGCAATCATATTTCG 120
Qy 239 CTGAGAGACAGAGAGAGGACATATCGGATATATAACCAAGCTTCGAGCCGGCAA 298
Db 121 CTGAGAGACAGAGAGAGGACATATCGGATATATAACCAAGTTTNGAGCCGGCAA 180
Qy 299 CGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCAC 358

Db 181 CGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGAGAGACTGTTTTCATGCTATTTCAN 240
Qy 359 TCTATTAAATCTTGCACTGCA 380
Db 241 TATATTAAATTTGCACTGCA 262

RESULT 3

US-08-721-489-4
; Sequence 4, Application US/08721489
; Patent No. 5786465
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallee, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Werberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,489
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-721-489-4

Query Match 47.6%; Score 238.4; DB 1; Length 279;
Best Local Similarity 92.4%; Pred. No. 8.2e-72;
Matches 242; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

Qy 119 ACCACATCCACTTTAAACAGCGGGCTTGCACTTACCTCACCTGACCAATCAGAGAG 178
Db 1 RCACATCCACCTTTAAACAGCGGGCTTGCAANAAGATNACCTTGACCAATCAGAGAG 60
Qy 179 CTCACATAATGCTAATTAGGCAAGAGAGGTAAAGAAATAGCCATCTATTTCG 238
Db 61 NTCANTAAATATATATTNGGCAAAAAGAGAGTAAAGAAATAGCAATCATATTTCG 120
Qy 239 CTGAGAGACAGAGAGAGGACATATCGGATATATAACCAAGCTTCGAGCCGGCAA 298
Db 121 CTGAGAGACAGAGAGAGGACATATCGGATATATAACCAAGTTTNGAGCCGGCAA 180
Qy 299 CGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCAC 358
Db 181 CGGCAACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTTCAN 240


```
RESULT 7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ19pt-F18
; US-08-232-463-14

Query Match 6.8%; Score 34; DB 1; Length 7218;
Best Local Similarity 7.1%; Pred. No. 0.3;
Matches 22; Conservative 155; Mismatches 135; Indels 0; Gaps 0;

QY 42 GAGATGGGGAGCTGAGAGACAGACTGATTTCTAGGCTGACTAAGATCCCTAA 101
DB 1332 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1273
QY 102 GCCTAGCTGGGAGGAGCAACATCCATTAAACGCGGGCTTGCACTTAGCTACA 161
DB 1272 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1213
QY 162 CCGACCAATCATGAGAGCTGACTAATGCTAATTAAGCAAGACAGAGTAAAGAAAT 221
DB 1212 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1153
QY 222 AGCCATCATCTATTGCTGAGAGCAACAGAGGAGGACAATGATCGGATATAAACA 281
DB 1152 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1093
QY 282 AGCTTCGAGCGGCAACGCAACCCCTTTGGTCCCTCCCTTTGTATGGAAGCTCG 341
DB 1092 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1033
```

```
QY 342 TTTTCATGCTAT 353
DB 1032 AATTAATCTGT 1021

RESULT 8
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
; GENERAL INFORMATION:
; APPLICANT: HERVE PERON
; APPLICANT: FREDERIC BESEME
; APPLICANT: FREDERIC BEDIN
; APPLICANT: GLAUCIA PARANHOS-BACCALA
; APPLICANT: FLORENCE KOMURIAN-PRADEL
; APPLICANT: COLETTE JOLIVET
; APPLICANT: BERNARD MANDRAND
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 38588
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1859 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-691-563C-46

Query Match 6.6%; Score 33; DB 3; Length 1859;
Best Local Similarity 6.92%; Pred. No. 0.32;
Matches 45; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 14 TCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGTGAAGACAGACTAGCTGG 73
DB 1687 TTCCCAAGACAGCTGGGGGTGTCCTTGAAGTGGGATTAAGAGTGAAGCAGCTGG 1746
QY 74 ATTTC 78
DB 1747 ACTTC 1751

RESULT 9
US-08-117-362-31/c
; Sequence 31, Application US/08117362
; Patent No. 5595872
; GENERAL INFORMATION:
; APPLICANT: WETTERAU II, John R.
; APPLICANT: SHARP, Daru Y.
; APPLICANT: GREGG, Richard E.
```

TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,362
FILING DATE: 03-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 783..890
US-08-117-362-31

Query Match 6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGAGCTGTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACCAATCCACC 130
Db 1345 TGCCCTTTTCAATGGGCTTAATCTCTTGAGGTCAATGCTAAGCAATTAAGAAA 1286
QY 131 TTTAAACACGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGCTCACTAAATG 190
Db 1285 TTTAAGAACTCGGTTTTTAAAAAAATCCTCAATTCAATGAAGAGCTGCATT 1226
QY 191 CTAATTAGGCAAGACAGAGAGTAAGAAATAGCCAATC---ATCTATGCTGAGAGC 246
Db 1225 CAATGTACCAACAACAGCAGTAACAATAGTGGTATCTTTTCTTTTAAAAAC 1166
QY 247 ACAGCAGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCCGGCAAGCAAC 306
Db 1165 AGACCAATGAATGAATAATTGAAGCAAGAACACATGCTGAGGAATGAATAAAGA 1106
QY 307 CCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTTCACTATTA 366
Db 1105 AATATCTGTTTTCTTAATGTTTGAAGGATTTGAATTTGTTTCATTTCAACTCAGCA 1046
QY 367 ATCTGCACTGCACTCTTCTGTGTCATGTTCTTACG 404
Db 1045 TGCTTTCACAAAACATTCCAAGTTACATGTTTGAAG 1008

RESULT 10
US-08-486-924-31/c
Sequence 31, Application US/08486924
GENERAL INFORMATION:
APPLICANT: Wetters II, John R.
APPLICANT: Sharp, Daru Y.

APPLICANT: Gregg, Richard E.
TITLE OF INVENTION: MICROSMAL TRIGLYCERIDE TRANSFER PROTEIN
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Burton Rodney
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543-4000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/486,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/117,362
FILING DATE: 03-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gaul, Timothy J.
REGISTRATION NUMBER: 33,111
REFERENCE/DOCKET NUMBER: DC21b
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252-5901
TELEFAX: (609) 252-4526
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1417 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 783..890
US-08-486-924-31

Query Match 6.4%; Score 32; DB 1; Length 1417;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 154; Conservative 0; Mismatches 180; Indels 4; Gaps 1;

QY 71 TGAATTCCTAGAGCTGTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACCAATCCACC 130
Db 1345 TGCCCTTTTCAATGGGCTTAATCTCTTGAGGTCAATGCTAAGCAATTAAGAAA 1286
QY 131 TTTAAACACGGGGCTTGCACTTAGCTACACCTGACCAATCAGAGCTCACTAAATG 190
Db 1285 TTTAAGAACTCGGTTTTTAAAAAAATCCTCAATTCAATGAAGAGCTGCATT 1226
QY 191 CTAATTAGGCAAGACAGAGAGTAAGAAATAGCCAATC---ATCTATGCTGAGAGC 246
Db 1225 CAATGTACCAACAACAGCAGTAACAATAGTGGTATCTTTTCTTTTAAAAAC 1166
QY 247 ACAGCAGAGGAGCAATGATCGGATATTAACCCAACTCTTGAGCCGGCAAGCAAC 306
Db 1165 AGACCAATGAATGAATAATTGAAGCAAGAACACATGCTGAGGAATGAATAAAGA 1106
QY 307 CCCTTTGGGTCCTCCCTTTGTATGGAGCTCTGTTTCAATGCTATTTCACTATTA 366
Db 1105 AATATCTGTTTTCTTAATGTTTGAAGGATTTGAATTTGTTTCATTTCAACTCAGCA 1046
QY 367 ATCTGCACTGCACTCTTCTGTGTCATGTTCTTACG 404
Db 1045 TGCTTTCACAAAACATTCCAAGTTACATGTTTGAAG 1008

RESULT 11
US-08-691-563C-58
Sequence 58, Application US/08691563C

Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Frederic BESME
APPLICANT: Frederic BEDI
APPLICANT: Glauca PARANHOS-BACCALA
APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELEPHONE: 703-836-2787
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1722 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-58

Query Match 6.4%; Score 32; IB 3; Length 1722;
Best Local Similarity 73.2%; Pred. No. 0.66;
Matches 41; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 GTCTCGGCCCACTCCCAACGACACTTGGTTTCCTGTGAGATGGGGACTGA 56
Db 1024 GTCTCGGCCCAATTCACGACGAGTGGGGTGTCTGTTTAGAGGGGGGATGA 1079

RESULT 12
US-08-973-273-1/C
Sequence 1, Application US/08973273
Patent No. 6140085
GENERAL INFORMATION:
APPLICANT: Dean, Caroline
APPLICANT: MacKnight, Richard C
APPLICANT: Bancroft, Ian
APPLICANT: Lister, Clare K
TITLE OF INVENTION: Genetic Control of Flowering
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye P.C.
STREET: 1100 No. 6140085th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/973,273
FILING DATE: 01-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/01332
FILING DATE: 03-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 951196.9
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-29
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9763 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Arabidopsis thaliana
STRAIN: Columbia
US-08-973-273-1

Query Match 6.1%; Score 30.8; DB 3; Length 9763;
Best Local Similarity 63.5%; Pred. No. 4.4;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 162 CCTGACCAATGAGAGCTCACTTAATGCTTAATTAGGCAAGACAGAGGTAAGAAT 221
Db 4845 CCTTCAGAAAGAAAGAACTATATACCTTGATAGGCAAGAAACAGGAAAGAAA 4786

Qy 222 AGCCAAATCATCTAT 235
Db 4785 GGATCAACATCCAT 4772

RESULT 13
US-09-739-455-3
Sequence 3, Application US/09739455
Patent No. 6413756
GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CLO00653
CURRENT APPLICATION NUMBER: US/09/739,455
CURRENT FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 23
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 11827
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(11827)
OTHER INFORMATION: n = A,T,C or G
US-09-739-455-3

Query Match 6.1%; Score 30.6; DB 4; Length 11827;
Best Local Similarity 51.1%; Pred. No. 5.8;
Matches 72; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY	103	CTTAGCTGGGAAGTGACCAACATCCACCCTTTAAACACGGGGCTTCCAACTTAGCTCAAC	162
Db	10853	CCCACTGCTGGAGTCGGAACTGCTGCCCTTTGTTGGGGCCCTTGTTCTTAAATCAGTT	10912
QY	163	CTGACCATTCAGAGGCTCATCAATTAATGCTATTATTTGGCAAAACACAGAGGTAAAGAAAT	222
Db	10913	CCCTTTAGAGTTTATTACATCAAAAAAAATTTAGTTTTTGAAGAAATTTGGAGAAAT	10972
QY	223	GCCAAATCATATTGGCTGAG	243
Db	10973	CAGAAACATGAATTTCAACGAG	10993

```

RESULT 14
US-09-422-576D-6
: Sequence 6, Application US/09422576D
: Patent No. 6395549
: GENERAL INFORMATION:
: APPLICANT: Tuan, Dorothy
: APPLICANT: Long, Qiaoming
: APPLICANT: Bengra, Chikn
: TITLE OF INVENTION: long Terminal Repeat, Enhancer, and Insulator Sequences for Use i
: TITLE OF INVENTION: Recombinant Vectors
: FILE REFERENCE: M0351-205010
: CURRENT APPLICATION NUMBER: US/09/422,576D
: CURRENT FILING DATE: 1999-10-21
: PRIOR APPLICATION NUMBER: US 60/105,256
: PRIOR FILING DATE: 1998-10-22
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6
: LENGTH: 1043
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-422-576D-6

```

	Query Match	Score	DB	Length
Best Local Similarity	62.1%	6.0%	4	1043
Matches	64	Conservative	0	Mismatches 38; Indels 1; Gaps 1;
Qy	253	GGAGGACACATATGCTGGGATATAAACCCAGCTTCGAGCGGCAACGGCAACCCCTTT	312	
Db	485	GCTGGGCGCCAGTAAAGAAATATAAAGCAGCGCTGCCGAGCCAGCACTGCAACCCGCG-TC	543	
Qy	313	GGGTCCCTCTCCCTTTGTATGAGAGCTCTGTTTTCATGCTATT	355	
Db	544	GGTCCCTTCACACTGTGAAGCTTGTGTTTCCTGCTCTTT	586	

```

RESULT 15
US-09-422-576D-5
: Sequence 5, Application US/09422576D
: Patent No. 6395549
:
: GENERAL INFORMATION:
: APPLICANT: Tuan, Dorochy
: APPLICANT: Long, Daoming
: APPLICANT: Bengia, Chikh
: TITLE OF INVENTION: Long Terminal Repeat, Enhancer, and Insulator Sequences for Use i
: TITLE OF INVENTION: Recombinant Vectors
: FILE REFERENCE: M0351-205010
: CURRENT APPLICATION NUMBER: US/09/422,576D
: PRIOR APPLICATION NUMBER: US 60/105,256
: PRIOR FILING DATE: 1998-10-22
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 5
:
: LENGTH: 1091
: TYPE: DNA
:
: ORGANISM: Artificial Sequence
:
: FEATURE:
:
: OTHER INFORMATION: Consensus Sequence

```

```
; NAME/KEY: misc feature
; LOCATION: (437)..(437)
; OTHER INFORMATION: n = any nucleotide
US-09-422-576D-5
```

Query Match	Best Local Similarity	Score	DB	Length
6.0%	62.1%	30.2	4	1091
Matches	64	Conservative	0	Mismatches 38; Indels 1; Gaps 1

Query	DB	Score	Length
533 GGGTCCCTTCACACCTGTGGAAGCTGTGTTCTTTCGCTCTTT	592	35.5	355
534 GGTGGGGCCAGTAAAGAAATTAAGACAGCGCTGCCCGACCGACAGTGGCAACCCGC	592	35.5	355
535 GAGAGGACATGATGGGATATAAACCAAGCTTCGACCGGCAACGCAACCCCTTT	312	35.5	312

Search completed: April 17, 2003, 07:39:12
Job time : 104.69 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:57:20 ; Search time 123.415 Seconds
(without alignments)
4077.979 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000
Perfect score: 501
Sequence: 1 gtctcgagcaaccctccca.....ctgagcatcagaggtgtcc 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 639749 seqs, 50280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCIT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCITUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	426	85.0	1329	7	US-08-979-847-108
2	375.4	74.9	2946	9	US-10-114-893-114
3	374.8	74.7	2930	10	US-09-902-535-1
4	341.8	68.2	635	7	US-08-979-847-102
5	317.6	63.4	180557	12	US-10-003-806-6
6	317.6	63.4	180557	12	US-10-003-806-9
7	311.4	62.2	410	10	US-09-880-107-538
8	305.2	60.9	326014	10	US-09-731-231A-3
9	286.6	57.2	541	10	US-09-864-761-8173
10	281.2	56.1	1894	10	US-09-864-761-8444
11	266.8	53.3	15425	9	US-10-091-504-1654
12	266.8	53.3	15425	9	US-09-764-869-1654
13	239.4	47.6	279	12	US-10-040-916-50
14	203.8	40.7	569	10	US-09-864-761-14951
15	146.2	29.2	246	10	US-09-864-761-20462
16	136.8	27.3	446	10	US-09-811-284-3
17	136.2	27.2	504	10	US-09-864-761-7027
18	129.6	25.9	3524	10	US-09-972-724-1
19	124.8	24.9	475	10	US-09-864-761-895

C 20	122.4	24.4	559	10	US-09-864-761-7501	Sequence 7501, Ap
C 21	111.6	22.3	440	10	US-09-864-761-3694	Sequence 3694, Ap
C 22	105.4	21.0	409	10	US-09-864-761-4153	Sequence 4153, Ap
C 23	92.2	18.4	579	10	US-09-864-761-13678	Sequence 13678, A
C 24	58.8	11.7	387	10	US-09-864-761-30194	Sequence 30194, A
C 25	58	11.6	525	10	US-09-893-737-31	Sequence 31, Appl
C 26	55.4	11.1	88	10	US-09-864-761-20907	Sequence 20907, A
27	50.6	10.1	275	10	US-09-864-761-17675	Sequence 17675, A
28	45.4	9.1	464	10	US-09-867-701-2947	Sequence 2947, Ap
29	42.6	8.5	929	9	US-10-093-766-14	Sequence 14, Appl
30	42.2	8.4	362	10	US-09-867-701-5411	Sequence 5411, Ap
C 31	41.8	8.3	1529	9	US-10-125-237-75	Sequence 75, Appl
C 32	41.2	8.2	81	10	US-09-864-761-24915	Sequence 24915, A
C 33	39.2	7.8	635	9	US-09-796-592-5087	Sequence 5087, Ap
34	39	7.8	3824	9	US-10-036-041-22	Sequence 22, Appl
35	39	7.8	3824	9	US-10-028-072-541	Sequence 541, Ap
36	39	7.8	3824	9	US-10-035-855-22	Sequence 22, Appl
37	39	7.8	3824	9	US-10-121-049-541	Sequence 541, Ap
38	39	7.8	3824	9	US-10-123-904-541	Sequence 541, Ap
39	39	7.8	3824	9	US-10-140-470-541	Sequence 541, Ap
40	39	7.8	3824	9	US-09-931-836-22	Sequence 22, Appl
41	39	7.8	3824	9	US-10-175-746-541	Sequence 541, Ap
42	39	7.8	3824	9	US-10-176-218-541	Sequence 541, Ap
43	39	7.8	3824	9	US-10-176-821-541	Sequence 541, Ap
44	39	7.8	3824	9	US-10-227-884-209	Sequence 209, Ap
45	39	7.8	3824	9	US-10-036-214-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-979-847-108
Sequence 108, Application US/08979847
Publication No. US2003003664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESIME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BRACALA, GLAUCIA
APPLICANT: KOMIRIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
NUMBER OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESS: OLIVE & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979, 847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 108:

SEQUENCE CHARACTERISTICS:
LENGTH: 1329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-108

Query Match 85.0%; Score 426; DB 7; Length 1329;
Best Local Similarity 90.6%; Pred. No. 8.2e-136;
Matches 453; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

2 TCTGGCCCACTCCCAACAGCACTTAGCTTTCTGTGAGATGGGGAGCTGAGAGAC 61
Db TCTCAGCCCACTCCCAACAGCACTTAGCTTTCTGTGAGAGGGTGAAGAGAGAC 556
QY 62 AGGACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACC 121
Db 557 AGGACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACC 616
QY 122 ACATCCACCTTTAAACAGCGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGAGCTC 181
Db 617 GCATCATCTTTTAAACATGGGGCTTGCACTTAGTTCACACCTGACCAATCAGAGAGCTC 676
QY 182 ACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATACCAATCATCTATTGCTTG 241
Db 677 ACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATACCAATCATCTATTGCTTG 736
QY 242 AAGACACAGCAGAGGAGCAATGATGGGATATTAACCAATCTTCAGCCGGCAACGG 301
Db 737 AAGACACAGCAGAGGAGCAATGATGGGATATTAACCAATCTTCAGCCGGCAACAG 796
QY 302 CAACCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCAATGCTATTTCACTCT 361
Db 797 CAACCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCACTATTTCACTCT 856
QY 362 ATTAATCTTGAACCTGACTCTTCTGTGCATGTTTCTTACGGCTTAGAGTGAAGCTTTC 421
Db 857 ATTAATCTTGAACCTGACTCTTCTGTGCATGTTTCTTACGGCTTAGAGTGAAGCTTTC 916
QY 422 GCTGGCATCCCACTGCTGTTTGGCCGACCGCAGACCCCGCTGACTCCCATCCCT 481
Db 917 GCTGGCATCCCACTGCTGTTTGGCCGACCGCAGACCCCGCTGACTCCCATCCCT 976
QY 482 CTGATCATGACGGGTGCTCC 501
Db 977 TTGATCCAGAGAGTGTCC 996

RESULT 2

US-10-114-893-134
Sequence 134, Application US/10114893
Publication No. US20020193567A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Collins-Racle, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Werberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Carlin-Duckett, McKeough
APPLICANT: Kelleher, Kerry S.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
FILE REFERENCE: GI 6000-10A
CURRENT APPLICATION NUMBER: US/10/114,893
CURRENT FILING DATE: 2002-04-02
EARLIER APPLICATION NUMBER: 09/413,232
EARLIER FILING DATE: 1999-10-06
NUMBER OF SEQ ID NOS: 321

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 134
LENGTH: 2946
TYPE: DNA
ORGANISM: Homo sapiens
US-10-114-893-134

Query Match 74.9%; Score 375.4; DB 9; Length 2946;
Best Local Similarity 99.7%; Pred. No. 3e-118;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

4 TCGGCCAACCCTCCCAACAGCACTTAGCTTTCTGTGAGATGGGGAGCTGAGAGAGAC 63
Db 2553 TCGGCCAACCCTCCCAACAGCACTTAGCTTTCTGTGAGATGGGGAGCTGAGAGAGAC 2612
QY 64 GACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACCAC 123
Db 2613 GACTAGCTGATTTCTTACGCTGACTAAGATCCCTTAAGCTTAGCTGGAAAGTGACCAC 2672
QY 124 ATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 183
Db 2673 ATCCACCTTTAAACAGCGGGCTTGCACTTAGCTCACACCTGACCAATCAGAGAGCTCAC 2732
QY 184 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 243
Db 2733 TAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCCATCATCTATTGCTGAG 2792
QY 244 AGCAGCAGAGGAGCAATGATGGGATATTAACCAAGCTTCGAGCCGGCAACGGCA 303
Db 2793 AGCAGCAGAGGAGCAATGATGGGATATTAACCAAGCTTCGAGCCGGCAACGGCA 2852
QY 304 ACCCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 363
Db 2853 ACCCCCTTTGGGTCCTCCCTTTGATGGAGCTGTGTTTCAATGCTATTTCACTCTAT 2912
QY 364 TAAATCTTGAACCTGCA 380
Db 2913 TAAATCTTGAACCTGCA 2929

RESULT 3

US-09-902-535-1
Sequence 1, Application US/09902535
Patent No. US20020102530A1
GENERAL INFORMATION:
APPLICANT: Keich, Jr., James C.
APPLICANT: McCoy, John M.
APPLICANT: Ml, Sha
TITLE OF INVENTION: Methods and compositions for diagnosing
TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
FILE REFERENCE: GIN-6006B4
CURRENT APPLICATION NUMBER: US/09/902,535
CURRENT FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: 60/216,657
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2930
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (930)...(2546)
US-09-902-535-1

Query Match 74.7%; Score 374.4; DB 10; Length 2930;
Best Local Similarity 99.7%; Pred. No. 6.6e-118;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
4 TCGGCCAACCCTCCCAACAGCACTTAGCTTTCTGTGAGATGGGGAGCTGAGAGAGAC 63

Db 2555 TCGGCCAACCTCCCAACAGCAGCTAGGTTTCTGTTGAGATGGGGGACGTGAGAGACAG 2614
Qy 64 GACTAGCTGATTTCTCTAGGCTGACTAGAAATCCCTAAGCCTAGCTGGGAAAGTGACCAC 123
Db 2615 GACTAGCTGATTTCTCTAGGCTGACTAGAAATCCCTAAGCCTAGCTGGGAAAGTGACCAC 2674
Qy 124 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCAGACCTTGACCAATCAGAGAGCTCAG 183
Db 2675 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCAGACCTTGACCAATCAGAGAGCTCAG 2734
Qy 184 TAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAG 243
Db 2735 TAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAG 2794
Qy 244 AGCAGCAGAGAGGAGCAGATGATCGGAGATATAACCCAAAGCTTGAGAGCCGCAACGGCA 303
Db 2795 AGCAGCAGAGAGGAGCAGATGATCGGAGATATAACCCAAAGCTTGAGAGCCGCAACGGCA 2854
Qy 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTGCTCTAT 363
Db 2855 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTGCTCTAT 2914
Qy 364 TAAATCTTGCAACTGC 379
Db 2915 TAAATCTTGCAACTGC 2930

RESULT 4
US-08-979-847-102
Sequence 102, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESSEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 635 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-102
Query Match 68.2%; Score 341.8; DB 12; Length 635;
Best Local Similarity 94.2%; Pred. No. 4.6e-107;
Matches 355; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 4 TCGGCCAACCTCCCAACAGCAGCTAGGTTTCTGTTGAGATGGGGGAGCTGAGAGACAG 63
Db 243 TCGGCCAACCTCCCAACAGCAGCTAGGTTTCTGTTGAGATGGGGGAGCTGAGAGACAG 302
Qy 64 GACTAGCTGATTTCTCTAGGCTGACTAGAAATCCCTAAGCCTAGCTGGGAAAGTGACCAC 123
Db 303 GACTAGCTGATTTCTCTAGGCTGACTAGAAATCCCTAAGCCTAGCTGGGAAAGTGACCAC 362
Qy 124 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCAGACCTTGACCAATCAGAGAGCTCAG 183
Db 363 ATCCACCTTTAAACACGGGGCTTGGCACTTAGCTCAGACCTTGACCAATCAGAGAGCTCAG 422
Qy 184 TAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAG 243
Db 423 TAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATAGCCAAATCATCTATTGCTGAG 482
Qy 244 AGCAGCAGAGAGGAGCAGATGATCGGAGATATAACCCAAAGCTTGAGAGCCGCAACGGCA 303
Db 483 AGCAGCAGAGAGGAGCAGATGATCGGAGATATAACCCAAAGCTTGAGAGCCGCAACGGCA 542
Qy 304 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTGCTCTAT 363
Db 543 ACCCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTTCATGCTATTGCTCTAT 602
Qy 364 TAAATCTTGCAACTGCA 380
Db 603 TAAATCTTGCAACTGCA 619

RESULT 5
US-10-003-806-6/c
Sequence 6, Application US/10003806
Patent No. US20020119929A1
GENERAL INFORMATION:
APPLICANT: Bishop, Colin E.
APPLICANT: Agoulnik, Alexander I.
APPLICANT: Zhu, Qichao
TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
FILE REFERENCE: P02066US1/10024824
CURRENT APPLICATION NUMBER: US/10/003,806
CURRENT FILING DATE: 2001-11-02
PRIOR APPLICATION NUMBER: 60/245,872
PRIOR FILING DATE: 2000-11-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 180557
TYPE: DNA
ORGANISM: Human
US-10-003-806-6
Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.9e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;
Qy 52 ACTGAGACAGAGACTAGCTGATTTCTCTAGGCTGACTAGAAATCCCTAAGCCTAGCTGG 111
Db 58991 AGTGAAGACAGAGACTAGCTGATTTCTCTAGGCTGACTAGAAATCCCTAAGCCTAGCTGG 58992
Qy 112 GAAAGTGACACATCACTTTAAACACGGGGCTTGGCACTTAGCTCAGACCTGACCAAT 171
Db 58991 GAAAGTGACACATCACTTTAAACACGGGGCTTGGCACTTAGCTCAGACCTGACCAAT 58992
Qy 172 C-----AGAGAGCTCACTAAATGCTAATTAGGCAAAAGCAGAGGTAAAGAAATA 222

Db 58871 CAGATAGTAAAGAGCTCACTAAATGCTAAITAGGCAAAAAAGAGAGTAAAGAAATA 58812
Qy 223 GCCATCATCTATTGCTGCTGAGAGACAGAGAGAGAGATGATGCGGATATATAACCCAA 282
Db 58811 GCCATCATCTATTGCTGCTGAGAGACAGAGAGAGAGATGATGCGGATATATAACCCAA 58752
Qy 283 GTCCTGAGCCGAGCAAGGCAACCCCTTTGGGTCCTCCCTTTGATATGAGAGCTCTGT 342
Db 58751 GCATCGAGCCAGCAACAGCTACGCTTTGGGTCCTCCCTTTGATATGAGAGCTCTGT 58692
Qy 343 TTTCATGCTATTCTCTATTAATCTTGCAATGCA--CTCTCTGCTCCATGTTTCT 400
Db 58691 -----CTTCACTATTATAATCTTGCAATGCACTGCACTCTCTTTGGTCTTACATTTGT 58642
Qy 401 TACGCTTGAAGTGAAGCTTTCGCTGCGCATCCACACTGCTGTTTGCAGCCAGCCAGAC 460
Db 58641 CATGCTGAGCTGAGCTTCTCTGCGCATCCACACTGCTGTTTGCAGCCAGCTGTG 58582
Qy 461 CCGCGCTGACTCCCATCCCTTGTGATGATGAGGCTGTG 500
Db 58581 CTGCTGCTGACTTCCATCCGTCAGATCCGCAAGAGTGC 58542

RESULT 6
US-10-003-806-9/c

; Sequence 9, Application US/10003806
; Patent No. US2002011929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulink, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003, 806
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245, 872
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-9

Query Match 63.4%; Score 317.6; DB 12; Length 180557;
Best Local Similarity 85.9%; Pred. No. 1.9e-97;
Matches 395; Conservative 0; Mismatches 44; Indels 21; Gaps 3;

Qy 52 ACTGAGAGACAGGACTAGCTGATTTCTTAGCTGCTAAGATCCCTAAGCTAGCTGG 111
Db 58991 ACTGAGAGACAGGACTAGCTGATTTCTTAGCTGCTAAGATCCCTAAGCTAGCTGG 58932
Qy 112 GAAGGTGACCACTCACTTTAAGACAGGGGCTTGCACTTAGCTCACACTGACCAAT 171
Db 58931 GAAGGTGACCGCTTCCACTTTAAACATGGGGCTTGCACTTAGCTCACACTGACCAAT 58872
Qy 172 C-----AGAGAGCTCACTAAATGCTAATTAG3CAAAAGACAGGAGTAAAGAAATA 222
Db 58871 CAGATAGTAAAGAGAGCTCACTAAATGCTAATTAG3CAAAAGACAGGAGTAAAGAAATA 58812
Qy 223 GCCATCATCTATTGCTGCTGAGAGACAGAGAGAGAGATGATGCGGATATATAACCCAA 282
Db 58811 GCCATCATCTATTGCTGCTGAGAGACAGAGAGAGAGATGATGCGGATATATAACCCAA 58752
Qy 283 GTCCTGAGCCGAGCAAGGCAACCCCTTTGGGTCCTCCCTTTGATATGAGAGCTCTGT 342
Db 58751 GCATCGAGCCAGCAACAGCTACGCTTTGGGTCCTCCCTTTGATATGAGAGCTCTGT 58692
Qy 343 TTTCATGCTATTCTCTATTAATCTTGCAATGCA--CTCTCTGCTCCATGTTTCT 400
Db 58691 -----CTTCACTATTATAATCTTGCAATGCACTGCACTCTCTTTGGTCTTACATTTGT 58642

Qy 401 TACGCTTGAAGTGAAGCTTTCGCTGCGCATCCACACTGCTGTTTGCAGCCAGCCAGAC 460
Db 58641 CATGCTGAGCTGAGCTTCTCTGCGCATCCACACTGCTGTTTGCAGCCAGCTGTG 58582
Qy 461 CCGCGCTGACTCCCATCCCTTGTGATGATGAGGCTGTG 500
Db 58581 CTGCTGCTGACTTCCATCCGTCAGATCCGCAAGAGTGC 58542

RESULT 7
US-09-880-107-538/c

; Sequence 538, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-NO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 538
; LENGTH: 410
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA250958
US-09-880-107-538

Query Match 62.2%; Score 311.4; DB 10; Length 410;
Best Local Similarity 92.6%; Pred. No. 1e-96;
Matches 349; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

Qy 4 TCGGCAACCTCCCAACAGACCTTAGGTTTCTGTTAGAGGGGAGCTGAGAGACAG 63
Db 377 TCAGCAACCTCCCAACAGACCTTAGGTTTCTGTTAGAGGGGAGCTGAGAGACAG 318
Qy 64 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGAGGAGTGACCCAC 123
Db 317 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGAGGAGTGACCCAC 258
Qy 124 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACTGACCAATCAGAGACTCAC 183
Db 257 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCACACTGACCAATCAGAGACTCAC 198
Qy 184 TAAATGCTAATTAGGCAAAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGTA- 242
Db 197 TAAATGCTAATTAGGCAAAAGACAGAGGTAAGAAATAGCAATCATCTATTGCTGTA- 138
Qy 243 GAGCAGACAGAGGAGCAATGATGCGGATTAACCCCAATCTTTCAGCCGCGC-AAAGG 301
Db 137 GAGCAGATGAGGAGGAGCAAGGATTTGCAATTAACCCCAATCTTTCAGCCAGCAAGG 78
Qy 302 CAACCCCTTTGGGTCCTCCCTTTGATAGGAGCTGTGTTTCATGCTATTCTACTCT 361
Db 77 CAACCCCTTTGGGTCCTCCCTTTGATAGGAGCTGTGTTTCATGCTATTCTACTCT 18
Qy 362 ATTAAATCTTTCAGACTG 378
Db 17 ATTAAATCTTTCAGACTG 1

RESULT 8
US-09-731-231A-3/c

; Sequence 3, Application US/09731231A
; Patent No. US20020082189A1

GENERAL INFORMATION:
APPLICANT: GUEBBER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: C0001007
CURRENT APPLICATION NUMBER: US/09/731,231A
CURRENT FILING DATE: 2000-12-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 326014
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(326014)
OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match Best Local Similarity 60.9%; Score 305.2; DB 10; Length 326014;
Matches 344; Conservative 0; Mismatches 33; Indels 9; Gaps 1;

Qy 4 TCGGCCAACCTCCCAACAGCACTTAGGTTTCTCTTGAATGGGGGACTGAGAGACAG 63
Db 170580 TCGGCCAACCTCCCAACAGCACTTAGGTTTCTCTTGAATGGGGGACTGAGAGACAG 170521
Qy 64 GACTAGCTGAGATTCTTCTAGCTGACTAAGATCCCTAAGCCTAGCTGGGAAGTGAACAC 123
Db 170520 GACTAGCTGAGATTCTTCTAGCTGACTAAGATCCCTAAGCCTAGCTGGGAAGTGAACAC 170461
Qy 124 ATCCACCTTTAAACAGCGGGGCTTGAACCTTAGCTACACCTGACCAATC-----AG 174
Db 170460 TTCTACCTTTAAACCGGGGCTTGAACCTTAGCTACACCTGACCAATGAGTAGGAAG 170401
Qy 175 AGAGCTCACTAAATGCTATTATAGCAAGACAGAGGTAAAGAAATAGCCATCATCTA 234
Db 170400 AGAGCTCACTAAATGCTATTATAGCAAGACAGAGGTAAAGAAATAGCCATCATCTA 170341
Qy 235 TTGCTGAGAGACAGACAGAGGAGCAATGATCGGATTTAAACCCAGTCTTCAGCCG 294
Db 170340 TTGCTGAGAGACAGACAGAGGAGCAATGATCGGATTTAAACCCAGTCTTCAGCCG 170281
Qy 295 GCAAGGGAACCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTCTGTTTCACTGATTT 354
Db 170280 GCAAGGGAACCCCTTTGGGTCCTCCCTCTTGTATGGGAGCTCTGTTTCACTGATTT 170221
Qy 355 TCACCTATTAAATCTTGCACTGCA 380
Db 170220 TCACCTATTAAATCTTGCACTGCA 170195

RESULT 9
US-09-864-761-8173/c
Sequence 8173, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmiga-X-1
CURRENT FILING DATE: 2001-05-23
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 8173
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC016663.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
US-09-864-761-8173

Query Match Best Local Similarity 57.2%; Score 286.6; DB 10; Length 541;
Matches 383; Conservative 0; Mismatches 59; Indels 21; Gaps 4;

Qy 49 GGGACTGAGAGACAGAGCTAGCTGATTTCTTGGCTGACTAAGATCCCTAAGCTTAGC 108
Db 536 GGGACTGAGAGACAGAGCTAGCTGATTTCTTGGCTGACTAAGATCCCTAAGCTTAGC 477
Qy 109 TGGGAAGGTGACCATCCTTAAACAGGGGCTTGCACTTAGCTACACCTGACC 168
Db 476 TGGGAAGGTGACCATCCTTAAACAGGGGCTTGCACTTAGCTACACCTGACC 417
Qy 169 AATC-----AGAGACTCACTAAATGCTAATTAGGAAAGACAGAGGTAAAGAA 219
Db 416 AATCAGTAGTAAGACAGCTCACATAAAGGCTAATTGGCTAAAGAGGTAAATTA 357
Qy 220 ATAGCAATCATCTATTGCTGAGAGACAGAGGAGGACATGATGGGATATTAACC 279
Db 356 ATAGCAATCATCTATTGCTGAGAGACAGAGGAGGATGATGATGGGATATTAACC 297
Qy 280 CAAGCTTCGAGCGGAGAGGCAACCCCTTTGGGTCCTCCCTTTGTTGGGAGCTC 339
Db 296 CAGGATTCAGCGAGAGGATGCAAGGCTTGAAGGTCCTCCCTCCACTGATGGAGCTC 237
Qy 340 TGTTCATGCTATTTCACCTATTAAATCTTGCAACTGACCTCT-TCGTGCTCATGTTT 398

Db 236 TGT-----TTTCACCTCTATTAACTCTGCACTGCACACTCTCTGATCTGTGTT 187
 QY 399 CTTACGGCTTAGACTGAGCTTTTGCTGCCCTGCACCACTGCTGTTTGCCGCACCGCAG 458
 Db 186 ATTCCGGTTTGAATTGAGCTTTTCTCTCACCATTCCACCACACTGCTATTACTGTCTGTGAG 127
 QY 459 AC-CGGCGCTGACTCCCATCCCTCTGATCATCAGAGGTTC 500
 Db 126 AACCGCTGCTGACTTCCACCCCTCCGAGATCTGCAGAGGTTC 84

```

RESULT 10
US-09-864-761-4444
: Sequence 4444, Application US/09864761
: Patent No. US20020048763N1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: CURRENT FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine v4rs. 1.1
: SEQ ID NO 4444
: LENGTH: 1894
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC002346.1
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4

```

?	OTHER INFORMATION:	EXPRESSED	IN BT174, SIGNAL = 5.9	
?	OTHER INFORMATION:	EXPRESSED	IN FETAL LIVER, SIGNAL = 4	
?	OTHER INFORMATION:	EXPRESSED	IN ADULT LIVER, SIGNAL = 6.7	
?	OTHER INFORMATION:	EXPRESSED	IN BRAIN, SIGNAL = 6.2	
?	OTHER INFORMATION:	EXPRESSED	IN BONE MARROW, SIGNAL = 6.3	
?	OTHER INFORMATION:	EXPRESSED	IN LUNG, SIGNAL = 3.9	
?	OTHER INFORMATION:	EXPRESSED	IN HELA, SIGNAL = 7.2	
US-09-864-761-4444				

Query Match	56.1%;	Score 281.2;	DB 10;	Length 1894;
Best Local Similarity	88.4%;	Pred. No. 5.4e-86;		
Matches 320;	Conservative 0;	Mismatches 33;	Indels 9;	Gaps 1;

OY	2	TCGAGGCACACTCCCCAA	CAGCACTAGATTTCTCGTGTGAGATGGGGAC	CTGAGAC	61
Db	1457	TGTTGGCCAACTCC	CCCAACAGACGTTGGATTTCTCTGTGAGAGGGGGAC	CTGAGAC	1516
OY	62	AGGACTAGCTGATTTCT	CTAGGCTGA	CTAAGATCCCTAAGCCTAGCTGGGAGGTGAC	121
Db	1517	AGGATTA	CTAGATTTCTTAGCA	CCAATTAAGATCCCTAAGACTAGCTGGGAGGTGAC	1576
OY	122	ACATCCACCTTTAAAC	ACGGGGCTTGCACTTAGCTCACACTGACCATC-----	172	
Db	1577	GCTTCCACCTTTAAAC	CCGGGCTTGCACTTAGCTCACGGCCCAACCATCAGATACAA	1636	
OY	173	AGAGAGCTACTAAATGCT	AATTTAGSCAAAGCAGGAGSTAAAGAAATAGCCATATC	232	
Db	1637	AGAGAGCTACTAAATGCT	AATTTAGSCAAAGCAGGAGATTAAGAAATAGCCATATC	1636	
OY	233	TATTTGCTGAGAGC	ACAGCAGGAGGACATGATCGGATATTAACCCAAAGTCTTGAGC	292	
Db	1697	TGTTGCTGAGAGC	ACAGCAGGAGGACATGATCGGATATTAACCCAGGATTCGAGC	1756	
OY	293	CGGCAAGGGCA	CCCCCTTTGGGTCCCCCTCCTTTGTATGSGAGCTCTGTTTCA	TGCTA	352
Db	1757	CAGCTCAGAGCTAC	CCCTCTTTGGGTCCCCCTCCTTTGTATGSGAGCTCTGTTTCA	CTCTA	1816
OY	353	TT	354		
Db	1817	TT	1818		

```

RESULT 11
US-10-091-504-1654/C
; Sequence 1654, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC00761
; CURRENT APPLICATION NUMBER: US/10/091,504
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2442
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1654
; LENGTH: 15425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-091-504-1654

```

Query Match	53.3%;	Score 266.8;	DB 9;	Length 15425;
Best Local Similarity	88.1%;	Pred. NO. 1.4e-80;		
Matches 317;	Conservative	0;	Mismatches 32;	Indels 11;
				Gaps 2;

0Y	4	TTGGCAACCTCCCAACACACACTGAGTTTCTCTTAGATGCGGACATAGAGACAG	63
Db	2049	TTGGCAACCTCCCAACACACACTGGGTTTTCTTTAGAGGGGGGACT--GAGACAG	1997
0Y	64	GACTGCTGATTTCTTCAAGCTGACTAGATCCCTAAACCTAGCTGGAAAGTACAC	123
Db	1991	AACCTGCTGATTTCTTCAAGTCCGACTAGATCCCTAAACCTAGCTGGAAAGTACTCC	1933

Qy	175	AGAGCTCACTAAATGCTATTATAGGCAAAAGACAGAGGTAAGAAATAGCCATCATCTA	234
Db	1871	AGAGCTCACAGAAATGCTAATTAGGCAAAAACAGAGGTAAACAAATAGCCATCATCTA	1812
Qy	235	TTTGCTTAGAGACACAGAGAGAGGAACAAATGTCGGGATTTAAACCCAACTCTTGAGCG	294
Db	1811	TCGCGTTAGAGACACAGTGGGAGGGAACAAATGTCGGGATTTAAACCCAGCATTTGAGCTG	1752
Qy	295	GCAACGGCAACCCCTTTGGGTCCTCCCTCTTTGTATGGAAGCTCTGTTTTCATGCTATT	354
Db	1751	GCAACAGTACCTCTTTGGGTCCCTCTCTTTGTATAGAGCTCTGTTTTCATGCTACT	1692

```

RESULT 12
US-09-764-869-1654/c
Sequence 1654, Application US/09764869
Patent No. US2002006151A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764,869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1654
LENGTH: 15425
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1654

```

Query Match	Similarity	Score	DB	Length
Best Local	Similarity	88.1%	Pred. No. 1,4e-80;	
Matches	317;	Conservative	0;	Mismatches 33; Indels 11; Gaps 2

QY	4	TCGGCCAACTCCCAACAGCACTTAGTGTTCCTGTGTGAGATGGGGGACGTAGAGACAG	63
Db	2049	TTGGCCAACTCCCAACAGCACTTAGTGTTCCTGTGTGAGAGGGGGGACT--GAGACAG	1992
QY	64	GACTGCGTGATTTTCCTTAGCGCTGACTTAAGAATCCCTTAACCTTAGCTGGGAAGTGAACC	123
Db	1991	AACTAGCTGGATTTTCCTTAGTCGACTTAAGAAATGCCCTTAACCTTAGCTGGGAAGTGAATGC	1932
QY	124	ATCCACTTTTAAACAGGGGCTTGCACCTTAGCTCAACCTACACATC-----AG	174
Db	1931	ATCCACTTTTAAACATGGGGCTTGCACCTTAGCTCAACCTCAACATAGGATTTAAAG	1872
QY	175	AGAGCTCACTAAATGCTAATTAGGCAAAAGACAGAGGTAAAGAAATAGCAATCATCTTA	234
Db	1871	AGAGCTCAACAGAAATGCTAATTAGGCAAAACAGAGGTAAACAAATAGCAATCATCTTA	1812
QY	235	TTGCTGTAGAGCAACAGCAGGAGGACAAATGATTCGGGATTTAAACCCAAAGCTTTCCAGCCG	294
Db	1811	TCGCTGTAGAGCAACGTGGAGGAGCAATGATTCGGGATTTAAACCCAGGCAATTTGAGCTG	1752
QY	295	GCAAGGGCAACCCCTTTGGGTCCTCCCTCTTGATATGGGAACCTGTTTCAATGCTATT	354
Db	1751	GCAAGGCTACCTCTTTGGGTCCTCCCTCTTTGTATAGAGTCTGTITTCACCTTACT	1692

RESULT 13
US-10-040-916-50
; Sequence 50, Application US/10040916
; Patent No. US20020146769A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallie, Edward
;

Racine, Lisa
 Merberg, David
 Treacy, Maurice
 Evans, Cheryl
 Spaulding, Vikki
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 ENCODING THEM
 NUMBER OF SEQUENCES: 71
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/040,916
 FILING DATE: 07-Jan-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/887,029
 FILING DATE: 07-FEB-1997
 APPLICATION NUMBER: 08/686,878
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Brown, Scott A.
 REGISTRATION NUMBER: 32,724
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8824
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 279 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 SEQUENCE DESCRIPTION: SEQ ID NO: 50:
 GS-10-040-916-50

Query Match	47.6%	Score 238.4	DB 12	Length 279
Best Local Similarity	92.4%	Pred. No. 9, 98-72		
Matches 242	Conservative 1	Mismatches 19	Indels 0	Gaps 0

QY	119	ACCA	CATC	CAC	CTTT	TAACA	CA	CGGG	CTTG	CA	CTTA	GTG	CT	CA	CC	TGA	CC	CA	TCA	GAG	178										
Db	1	RC	CA	CA	CT	CTTT	TAACA	CA	CGGG	NTTG	CA	AA	NA	AG	TTNA	CA	CTTG	CA	CA	TCA	GAG	60									
QY	179	CT	CA	CT	AA	AT	TG	CT	PA	TT	AG	SC	AA	NA	GA	CA	GA	GG	TA	AA	GA	AA	TA	AG	CC	CA	TAT	CT	ATT	GC	238
Db	61	NT	CA	NT	AA	AT	TG	AT	TA	TT	AG	SC	AA	NA	GA	CA	GA	GG	TA	AA	GA	AA	TA	AG	CC	CA	TAT	CT	ATT	GC	120
QY	239	CT	GA	GA	CA	CA	CG	CA	GG	AG	GA	CA	AT	GA	TG	CG	GA	TAT	AA	CC	CA	AG	CT	TT	CG	AG	CC	GG	CA	298	
Db	121	CT	GA	GA	CA	CA	CG	CA	GG	AG	GA	CA	AT	GA	TG	CG	GA	TAT	AA	CC	CA	AG	CT	TT	CG	AG	CC	GG	CA	180	
QY	299	CG	GA	CA	CCCC	CTTT	GG	CT	CCCC	CT	CC	TTT	GA	TG	GA	TG	GA	AG	CT	CG	TTT	TAT	AT	GT	TAC	358					
Db	181	CG	GA	CA	CCCC	CTTT	GG	CT	CCCC	CT	CC	TTT	GA	TG	GA	TG	GA	AG	CT	CG	TTT	TAT	AT	GT	TAC	240					
QY	359	TCT	AT	TA	AT	CT	TG	CA	CT	GA	TG	CA	380																		
Db	241	TNT	AT	TA	AT	NT	TG	CA	CT	GA	TG	CA	262																		

RESULT 14
US-09-864-761-14951/c
; Sequence 14951, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wenhang

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263,6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 14951

LENGTH: 569

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AP000233.1

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9.3

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1

US-09-864-761-14951

Query Match 40.7%; Score 203.8; DB 10; Length 569;

Best Local Similarity 79.3%; Pred. No. 1.1e-59;

Matches 337; Conservative 0; Mismatch 62; Indels 26; Gaps 7;

52 ACTGAGACAGAGACTAGCTGATTTCTTAGCTGACTAAGATCCCTAAGCCTGAGTGG 111

DB 415 ACCGAGACAGAGACTAGCTGATTTCTTAGCCGACCTAAGATTCCTAAGTCTAGCTGG 356

412 G-AAGGTACACACATCCACTTTAAACAGGGGGCTGCACTTAGCTGCACACCTGACCAA 170

335 GAAAGGTACACACACCTTTAACAACGAGAGCTGTAACTTAGCTGCATCCGACCAA 236

Qy	171	TC-----AGGAGCTCACTAAATGCTAATTTAGGC--AAAGACAGGAGGTAAACAA	220
Db	295	TCAGGTAGTAAAGGGGTTCACGTAAATTACAAATTTAGGCTTAAAGCAGGAGGTAAAGAA	230
Qy	221	TAG--CCATCATCTATTGCTCTGAGAGCACAGACAGGAGGACAAATGATCCGGATATTAACC	279
Db	235	TATGTCAATTCATATATGCTTACGCTTACAGCACAGGGGAGAGGCATATGATTTGGGATATTAAC	176
Qy	280	CAA--GTCTTCGAGCCGGCAACGGCAACCCCTTTGGGTCCCTCCCTTGTGTATGGAGC	333
Db	175	CCAGGAGATTCAACCGGGAGTGGGCAACCCCTTTGGGTCCCTCCCATTTGATGGAGC	116
Qy	338	TCCTGTTTCATGCTATTCTACTCTATTAAATCTTGCAACTG--CACTCTTCTGTCCATG	393
Db	115	TCCTG-----TTTCACTCTGTATAAATCTTGCAACTGTACACTCTCTGTGCTAGTG	66
Qy	336	TTTCTTAGGGCTTAGAGCTGAGCTTTCCTCGCCATCCACCACTGTCGTCTTTGGCGGCCACCG	455
Db	65	TTTGTTCGGGCTCAAGCTGAACCTTTTGCTCACCTCTACCACTGCTGTCTTCCCTGTAG	6
Qy	456	CAGAC 460	
Db	5	CAGAC 1	

```

RESULT 15
US-09-864-761-20462/c
Sequence 20462, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmica-X-1
CURRENT FILING DATE: 2001-05-23
PRIORITY FILING DATE: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30

```

```

; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20462
; LENGTH: 246
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010951.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
; OTHER INFORMATION: NT HIT: AB026898.1, EVALUE 4.00e-57
; OTHER INFORMATION: EST HUMAN HIT: A1492055.1, EVALUE 3.00e-49
; OTHER INFORMATION: SWISSPROT HIT: Q02279, EVALUE 7.30e-01
US-09-864-761-20462

```

```

Query Match      29.2%; Score 146.2; DB 10; Length 246;
Best Local Similarity 88.5%; Pred. No. 4.4e-40;
Matches 185; Conservative 0; Mismatches 13; Indels 11; Gaps 2;

QY 232 CTATTGCTGAGAGACAGCAGAGGAGCAATGATCGGATATTAACCAAGCTTCGAG 291
   |||||||
Db 246 CTGTTGCTGAGAGACAGCAGGAGGACAATTATCAGGATATAAACCAAGCATTGAG 187
   |||||||

QY 292 CCGGACAGGACACCCCTTTGGGATCCCTCCCTTGTATGGGAGCTCTGTTTCATGCT 351
   |||||||
Db 186 CTGGCAAGGTAAACCCCTTTGGGATCCCTCCCTTGTATGGGAGCTCTAT----- 136
   |||||||

QY 352 ATTCACTCTAATTAATCTTGCACTGCTCTTCTGTCATGTTTCTTACGGCTTGAG 411
   |||||||
Db 135 -CTTCACTCTAATTAATCTTGCACTGCTCTTCTGTCATGTTTCTTAC-GCTTGAG 78
   |||||||

QY 412 CTGAGCTTTCGCTGCGCATCCACCACTGC 440
   |||||||
Db 77 CTGAGCTTTCGCTGCGCATCCACCACTGC 49
   |||||||

```

Search completed: April 17, 2003, 07:49:41
 Job time : 300.415 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:30:25 ; Search time 1268.84 Seconds

(without alignments)
6394.748 Million cell updates/sec

Title: US-09-719-554-3_COPY_9500_10000

Perfect score: 501

Sequence: 1 gtcctgcgcacactcccca.....ctgcatcatgcagcgtctcc 501

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	376	75.0	440	12	BE732673
2	372.8	74.4	494	9	AA781423
3	372.2	74.3	436	9	AI128526
4	365.8	73.0	415	9	AI128496
5	360.4	71.9	388	14	H01325
6	355.4	70.9	727	17	AG029908

C	7	348.4	69.5	422	14	NS3177	NS3177 yv56h11.s1
	8	345.2	68.9	771	13	BI087886	BI087886 602852690
	9	343.8	68.6	683	17	AG102951	AG102951 Pan trogl
C	10	339.6	67.8	385	14	NS5091	NS5091 yv43e03.s1
	11	337.8	67.4	458	14	R76086	R76086 y171b03.s1
C	12	332.4	66.3	522	10	AW971553	AW971553 EST383642
	13	329	65.7	353	14	T69704	T69704 yd13a03.s1
C	14	326.6	65.2	342	9	AA860368	AA860368 aj59c05.s
C	15	326.4	65.1	424	14	R27412	R27412 yh46d11.s1
C	16	324.4	64.8	342	14	T47345	T47345 yb10h02.s1
C	17	318.8	63.6	438	14	R77278	R77278 y175d06.s1
C	18	317.6	63.4	443	9	AA837267	AA837267 od26b10.s
C	19	312.6	62.4	433	9	AI379210	AI379210 cd01g11.x
C	20	312.6	62.4	490	9	AI598135	AI598135 tm14a10.x
C	21	311.4	62.2	410	9	AA250958	AA250958 zso7d10.s
C	22	311	62.1	425	9	AI570707	AI570707 tm79g09.x
C	23	309.4	61.8	446	9	AI393478	AI393478 tg45g04.x
C	24	307.2	61.3	921	17	BH149565	BH149565 ENTQ48TR
C	25	305.4	61.0	463	14	R68685	R68685 y114g06.s1
C	26	302.6	60.4	404	14	R27389	R27389 yh46a09.s1
C	27	296.2	59.1	328	9	AA729556	AA729556 nx58c05.s
C	28	293.6	58.6	485	10	AW511366	AW511366 hd45h03.x
C	29	293	58.5	431	9	AA552941	AA552941 nk61a10.s
C	30	290.8	58.0	471	9	AA709471	AA709471 zf19h06.s
C	31	289.6	57.8	653	17	AG033781	AG033781 Pan trogl
C	32	287.8	57.4	701	17	AG126669	AG126669 Pan trogl
C	33	285	56.9	611	17	AQ381711	AQ381711 RPCI11-16
C	34	284	56.7	712	17	AO892947	AO892947 HS 3131.B
C	35	281.2	56.1	722	17	AG049481	AG049481 Pan trogl
	36	279.2	55.7	503	14	BM723292	BM723292 UI-E-EUO-
C	37	279.2	55.7	530	14	BM682297	BM682297 UI-E-EUO-
C	38	278.4	55.6	470	9	AI074704	AI074704 ox83d05.s
C	39	277.6	55.4	446	9	AI288235	AI288235 q18a08.x
C	40	276.2	55.1	960	17	AO900343	AO900343 HS 3179.B
C	41	273.2	54.5	893	14	BQ437925	BQ437925 ASENCOURT
C	42	272.8	54.5	609	17	AG066901	AG066901 Pan trogl
C	43	272.4	54.4	326	14	D29167	D29167 HUMNR203 Hu
	44	272.4	54.4	447	12	BF919416	BF919416 QVO-NT015
	45	271.8	54.3	681	10	AV722664	AV722664 AV722664

ALIGNMENTS

RESULT 1
BE732673
LOCUS
DEFINITION
601571305F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3925728 5',
mRNA sequence.
ACCESSION
BE732673
VERSION
BE732673.1 GI:10146665
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 440)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Straubeberg, Ph.D.
Email: cgabds-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: L1CM752 row: j column: 01
High quality sequence stop: 440.
Location/Qualifiers
1..440
/organism="Homo sapiens"

FEATURES

source

/db_xref="taxon:9606"
/clone="IMAGE:3925728"
/clone_1id="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pOTB7; Site: 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: 3GCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and SuperScript II RT (Life Technologies)."

BASE COUNT 124 a 117 c 101 g 98 t

ORIGIN

Query Match 75.0%; Score 376; DB 12; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.8e-111;
Matches 376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 65 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 124
OY 64 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACCAC 123
DB 125 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACCAC 184
OY 124 ATCCACCTTTAAACACGGGGCTTGCACCTAGCTACACCTGACCAATCAGAGACTCAC 183
DB 185 ATCCACCTTTAAACACGGGGCTTGCACCTAGCTACACCTGACCAATCAGAGACTCAC 244
OY 184 TAAATGCTAATTAGGCAAGAGAGAGTAAAGTAATGCAATCATCTATTGCTGAG 243
DB 245 TAAATGCTAATTAGGCAAGAGAGAGTAAAGTAATGCAATCATCTATTGCTGAG 304
OY 244 AGCAGCAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGAGCGGCAACGGCA 303
DB 305 AGCAGCAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGAGCGGCAACGGCA 364
OY 304 ACCCCCTTTGGGTCCCTCTTGTATGGAGCTTGTCTTCATGCTATTCTACTCTAT 363
DB 365 ACCCCCTTTGGGTCCCTCTTGTATGGAGCTTGTCTTCATGCTATTCTACTCTAT 424
OY 364 TAAATCTTGCACTGC 379
DB 425 TAAATCTTGCACTGC 440

RESULT 2
AA781423 494 bp mRNA linear EST 31-DEC-1998
LOCUS a36c03.81 Soares_testis_NHT Homo sapiens cDNA clone 1391428 3'
DEFINITION similar to contains PTR7.1 PTR7 repetitive element ;, mRNA
sequence.

ACCESSION AA781423

VERSION AA781423.1 GI:2840754

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-1email.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

cDNA Library Arrayed by: Greg Lemmon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.1nl.gov/bbrp/image/image.html
Insert Length: 1645 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 475.
Location/Qualifiers

FEATURES

SOURCE

1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1391428"
/clone_1id="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACCAATCTAGAGGAGGCGGCCCAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 111 c 133 g 141 t 1 others

ORIGIN

Query Match 74.4%; Score 372.8; DB 9; Length 494;
Best Local Similarity 99.2%; Pred. No. 2.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 63
DB 382 TCGGCCAACCTCCCAACAGCACTAGGTTTCTGTTGAGATGGGGAGCTGAGAGACAG 323
OY 64 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACCAC 123
DB 322 GACTAGCTGATTTCTGAGCTGACTAAGATCCCTAAGCTAGCTGGAGAGTACCAC 263
OY 124 ATCCACCTTTAAACACGGGGCTTGCACCTAGCTACACCTGACCAATCAGAGACTCAC 183
DB 262 ATCCACCTTTAAACACGGGGCTTGCACCTAGCTACACCTGACCAATCAGAGACTCAC 203
OY 262 TAAATGCTAATTAGGCAAGAGAGTAAAGTAATGCAATCATCTATTGCTGAG 243
DB 202 TAAATGCTAATTAGGCAAGAGAGTAAAGTAATGCAATCATCTATTGCTGAG 143
OY 244 AGCAGCAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGAGCGGCAACGGCA 303
DB 142 AGCAGCAGAGAGGAGCAATGATCGGATATAACCCAGTCTTGAGCGGCAACGGCA 83
OY 304 ACCCCCTTTGGGTCCCTCTTGTATGGAGCTTGTCTTCATGCTATTCTACTCTAT 363
DB 82 ACCCCCTTTGGGTCCCTCTTGTATGGAGCTTGTCTTCATGCTATTCTACTCTAT 23
OY 364 TAAATCTTGCACTGC 380
DB 22 TAAATCTTGCACTGC 6

RESULT 3
A1128526 436 bp mRNA linear EST 27-OCT-1998
LOCUS A1128526

DEFINITION g66h10.x1 Soares_placenta_8c05week_2bH8b09M Homo sapiens cDNA

clone IMAGE:1724147 3' similar to contains PTR7.1 PTR5 repetitive

element ;, mRNA sequence.

ACCESSION A1128526

VERSION A1128526.1 GI:3597040

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 436)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 720 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence, stop: 428.
Location/Qualifiers
1. .436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_placenta_8to9weeks_2NBH8to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Faciina Bernaldo."

BASE COUNT 95 a 96 c 106 g 139 t
ORIGIN
Query Match 74.3%; Score 372.2; DB 9; Length 436;
Best Local Similarity 99.2%; Pred. No. 3.1e-110;
Matches 374; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 TCGGCAACTCCCAAGACGACTAGGTTTCTGTTGAGATGGGGAGTGAAGACAG 63
DB 399 TCGGCAACTCCCAAGACGACTAGGTTTCTGTTGAGATGGGGAGTGAAGACAG 340
QY 64 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGGTGACAC 123
DB 339 GACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGGTGACAC 280
QY 124 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTACCACTGACCAATCAGAGACTCAC 183
DB 279 ATCCACCTTTAAACAGGGGGCTTGCACTTACCTACCACTGACCAATCAGAGACTCAC 220
QY 184 TAAATGCTAATTAGGCAAGAAGAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 243
DB 219 TAAATGCTAATTAGGCAAGAAGAGAGGTAAGAAATAGCAATCATCTATTGCTGAG 160
QY 244 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 303
DB 159 AGCAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACGCA 100
QY 304 ACCCTTTGGGTCCCTTCCTTTGTATAGGAGCTCTGTTTCAATGCTATTCACTCTAT 363
DB 99 ACCCTTTGGGTCCCTTCCTTTGTATAGGAGCTCTGTTTCAATGCTATTCACTCTAT 40
QY 364 TAAATCTTGCAACTGCA 380
DB 39 TAAATCTTGCAACTGCA 23

RESULT 4
LOCUS A1128496 415 bp mRNA linear EST 27-OCT-1998
DEFINITION qc1e08.x1 Soares_placenta_8to9weeks_2NBH8to9W Homo sapiens cDNA
clone IMAGE:1714118 3' similar to contains PTR.tl PTR7 repetitive
element //, mRNA sequence.
ACCESSION A1128496

VERSION A1128496.1 GI:3597010
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 415)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1184 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence, stop: 413.
Location/Qualifiers
1. .415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_placenta_8to9weeks_2NBH8to9W"
/dev_stage="two placentae: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Faciina Bernaldo."

BASE COUNT 94 a 96 c 107 g 116 t 2 others
ORIGIN
Query Match 73.0%; Score 365.8; DB 9; Length 415;
Best Local Similarity 97.6%; Pred. No. 3.7e-108;
Matches 370; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GTCTCGCAACTCCCAAGACGACTAGGTTTCTGTTGAGATGGGGAGTGAAGAGA 60
DB 379 GTCTCGCAACTCCCAAGACGACTAGGTTTCTGTTGAGATGGGGAGTGAAGAGA 320
QY 61 CAGACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGGTGAC 120
DB 319 CAGACTAGCTGATTTCTTGAAGCTGACTAAGATCCCTAAGCTAGCTGGAGGTGAC 260
QY 121 CACATCCACCTTTAAACAGGGGGCTTGCACTTACCTACCACTGACCAATCAGAGACT 180
DB 259 CACATCCACCTTTAAACAGGGGGCTTGCACTTACCTACCACTGACCAATCAGAGACT 200
QY 181 CACTAATGCTAATTAGGCAAGAAGAGAGGTAAGAAATAGCAATCATCTATTGCTG 240
DB 199 CACTAATGCTAATTAGGCAAGAAGAGAGGTAAGAAATAGCAATCATCTATTGCTG 140
QY 241 GAGAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACG 300
DB 139 GAGAGCAGAGAGGAGCAATGATCGGATATTAACCAAGTCTTGAGCCGGCAACG 80
QY 301 GCAACCCCTTTGGGTCCCTTCCTTTGTATAGGAGCTCTGTTTCAATGCTATTCACTC 360
DB 79 GCAACCCCTTTGGGTCCCTTCCTTTGTATAGGAGCTCTGTTTCAATGCTATTCACTC 20
QY 361 TATTAAATCTTGCAACTGC 379
DB 19 TATTAAATCTTGCAACTGC 1

RESULT 5

LOCUS/c	H01325	388 bp	mRNA	linear	EST 19-JUN-1995
DEFINITION	Y199E01.s1 Soares placenta NBZHP Homo sapiens CDNA clone				
ACCESSION	IMAGE:147384 3', mRNA sequence.				
VERSION	H01325				
KEYWORDS	H01325.1	GI:864258			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 388)				
AUTHORS	Hillier,L., Clark,N., Dubucq,T., Elliston,K., Hawkins,M., Holman, M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marras,M., Parsons,J., Riklin,L., ROLLING,I., Soares,H., Tan,F., Trevaaskis,E., Waterson, R., Williamson,A., Wohlmann,P. and Wilson,R. The WashU-Merck EST Project Unpublished (1995) Contact: Wilson RK Washington University School of Medicine 444 Forest Park Parkway, Box 850L, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 790				
TITLE	High quality sequence strops. 346				
JOURNAL	Source: IMAGE Consortium, LLNL				
COMMENT	This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 790 Std Error: 0.00 Seq primer: Promega -21ml3 High quality sequence strop. 346. Location/Qualifiers 1..388				
FEATURES					
source	1..388				

```

/organism="Homo sapiens"
/db_xref="GDB:559031"
/db_xref="Caxon:9606"
/clone_image="147384"
/clone_lib="Soares placenta NB2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/notes="Host: placenta; Vector: pTR73 (pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AAGTGAAGAATTGCGGCCGCGAGAAATTTTATTTTATTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTR73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Patima Ronaldo. "
```

Query Match	71.9%	Score 360.4	DB 14	Length 388
Best Local Similarity	97.8%	Pred. No. 2,1e-106		
Matches 364	Conservative	0	Mismatches 8	Indels 0
				Gaps 0
QY	4	TTGGCCAACTCCCAACGCACTTATGGTTTCCTGTGAGATGGGGGACTAGAGACAG	63	
DB	377	TCGGCAACCTCCCAANNAGCACTTAGTGTTCCTGTGAGATGGGGGACTAGAGACAG	318	
QY	64	GACTAGCTGGATTTCTCTAGCGCTGACTAAGATCCCTTAAGCTTGGGAAGGTGACAC	123	
DB	317	GACTAGCTGGATTTCTCTAGCGCTGACTAAGATCCCTTAAGCTTGGGAAGGTGACAC	258	
QY	124	ATCCACCTTTAAACAGGGGCTTGCAACTAGCTCAACACTGACCAATCAGAGGTCAAC	183	
DB	257	ATCCACCTTTAAACAGGGGCTTGCAACTAGCTCAACACTGACCAATCAGAGGTCAAC	198	
QY	184	TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCCATATCTATTGCTTGAG	243	
DB	197	TAAATGCTAATTAGGCAAGACAGAGGTAAGAAATAGCCATATCTATTGCTTGAG	138	

QY	DB	QY	DB
244	137	304	77
AGCACAGAGAGGACAATGATGGGATTTAAACCAAGCTTCGAGCGGCAACGGCA	AGCACAAGCAGGAGGACAAATATGGGATTTAAACCAAGCTTCGAGCGGCAACGGCA	AGCCCTTTGGGATCCCTCCCTTGATGGGAGCTGTGTTTCATGCTATTTCACTAT	AGCCCTTTGGGATCCCTCCCTTGATGGGAGCTGTGTTTCATGCTATTTCACTAT
303	78	363	18
QY	DB	QY	DB
364	17	364	17
TAATCTTGCAA	TAATCTTGCAA	TAATCTTGCAA	TAATCTTGCAA
375	6	375	6
QY	DB	QY	DB
17	6	17	6
TAATCTTGCAA	TAATCTTGCAA	TAATCTTGCAA	TAATCTTGCAA
6	6	6	6

RESULT 6	
AC029908	
LOCUS	AC029908 727 bp DNA linear GSS 01-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-002C04.R, genomic survey sequence.
ACCESSION	AC029908
VERSION	AC029908.1 GI:1656781
KEYWORDS	GSS.
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male

ORGANISM	BAC Library clone:PTB-002C04.R.
Pan troglodytes	
Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:	
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Pan.	
REFERENCE	
AUTHORS	1
	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
TITLE	Totoki, Y., Watanabe, H. and Sakaki, Y.
JOURNAL	BAC end sequences of library PTB
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 727)
	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
	Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Aao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suchoi-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimbes@gsc.riken.go.jp, URL: <http://bgp.gsc.riken.go.jp/>, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PB3 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

```

PRIMERS
Sequencing: M13rev
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI.
FEATURES
Location/Qualifiers
1..727
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-002C04.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT
170 a    215 c    175 g    166 t    1 others
ORIGIN

```

Query Match	70.9%	Score 355.4	DB 17	Length 727
Best Local Similarity	91.1%	Pred. No. 1.2e-104		
Matches 411	Conservative 0	Mismatches 36	Indels 4	Gaps 3
QY 52	ACTGAGACAGAGATTAAGTGGATTTTCTAGAGCTACATTAAGATTCCTTAAGCTTAGCGG	111		
DB 275	ATTGAGAGACAGAGATTAAGTGGATTTTCTAGAGCTACATTAAGATTCCTTAAGCTTAGCGG	334		
QY 112	GAGGTGACACATCCAGCTTTAAACAGGGGCTTGCAACTTAGCTCAACCTGACAT	171		
DB 335	GAGGTGACAGGATCCACTTTAAACAGGGGCTTGCAACTTTGCTCAACCCAACTCAT	394		
QY 172	CAGAGACTCAATAAATGCTTAATTAGGCAAAAGACAGAGTAAAGATAATGCCATCAT	231		

Db	395	CGGAGGCTCATTAATAATGCTTAATTAGGCATAAAGAGAGGTAAAGAAATAGCCATAT	454
Qy	232	CTATTGCTGTGAGCAGCAGCAGAGAGGGAACAATGATCGGATATTAACCAGTCTTGAG	291
Db	455	CTATTGCTGTGAGAGCAGTAGTGGGAAGATTAAGATTCGGGATATAAACCGAGCATTTCAAG	514
Qy	292	CCGGAAAGGGAAACCCCTTTTGGGTCCCTCCCTTTGTATAGGAAGCTGTGTTTATAGT	351
Db	515	CGAGAAAGGGCAACCCCTTTTGGGT - CCTCCCTTGTATAGGAAGCTGTGTTTACTCT	573
Qy	352	ATTTCATCTTAATTAATCTTGCAACTGCACTCTTGTGTCATGTTCTTAACGCTTGAG	411
Db	574	ATTTCATCTTAATTAATCTTGCAACTGCACTCTTGTGTCGTTGTTGTACGGCTCGAG	633
Qy	412	CTGAGCTTTCGTCGCCATTCACCACTGCTGTT - TGCCGCCACCGGAGACCCGCCGCTGA	470
Db	634	CTGAGCTTTCCTCCGCCATTCACCACTGCTGTTTTCGCCGCCATTCGACGCCGCTTGA	693
Qy	471	CTCCCATCCCTCTGGATCATGAGGAGTGGCC	501
Db	694	CTTCATCCCTTC - GATCCAGCAGGAGTGTCC	722

LOCUS	422 bp	mRNA	linear	EST 28-JAN-1997
DEFINITION				
	N53177			
	iy56h11.s1	Soares fetal liver spleen	INFLS Homo sapiens	cDNA clone
	IMAGE:246789	3' similar to contains	PTR.t2	PCR7 repetitive element
	;;	mRNA sequence.		

ACCESSION	N53177	GI:1194343
VERSION	N53177.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 422)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

TITLE
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (3), 807-828 (1996)

MEDLINE 97044478
 COMMENT Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

Insert length: 938 Std Error:
Seq primer: m13 -40 forward
High quality sequence stop: 301.

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="GDB:3796035"
/db_xref="taxon:9606"
/clone="IMAGE:246789"
/clone_1ib="Soares fetal liver spleen INFLs"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="MDH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) p:meae
15' AACTGGAAGAATTATTAAGAAGCTTTTTTTTTTTTTTTT 3'),
double-stranded cDNA was ligated to Eco RI adaptors

```

(Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

Query Match	69.5%;	Score 348.4;	DB 14;	Length 422;
Best Local Similarity	98.7%;	Pred. No. 1.8e-102;		
Matches 372;	Conservative	0;	Mismatches 2;	Indels 3; Gaps 2;

Qy 4 TCGGCAACCTCCCAACAGACTTAGSTTTTCCTGTGATGGGGGACTGAGAGACAG 63
|||||
Db 375 TCGGCAACCTCCCAACAGACTTAGSTTTTCCTGTGATGGGGGACTGAGAGACAG 316

Oy	64	GACTGCTGGATTTCCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACAC	123
Db	315	GACTGCTGGATTTCCTAGGCTGACTAAGAAATCCCTAAGCCTAGCTGGAAAGTGACAC	256

Oy	124 ATCCACCTTTAAACACGGGGGCTTGAATTAGCTCACCGCAACAATCAGAGCTCAC	183
Dd	255 ATCCACTTTTAAACACGGGGGCTTGGAATTAGCTCACCGCAACAATCAGAGCTCAC	196

303	AGCCATGAGAGGAGCAATGATGCGGATATTAATACCCCAAGCTTTCGAGCTGGGCAAGGCA
304	
305	
306	
307	
308	
309	
310	
311	
312	
313	
314	
315	
316	
317	
318	
319	
320	
321	
322	
323	
324	
325	
326	
327	
328	
329	
330	
331	
332	
333	
334	
335	
336	
337	
338	
339	
340	
341	
342	
343	
344	
345	
346	
347	
348	
349	
350	
351	
352	
353	
354	
355	
356	
357	
358	
359	
360	
361	
362	
363	
364	
365	
366	
367	
368	
369	
370	
371	
372	
373	
374	
375	
376	
377	
378	
379	
380	
381	
382	
383	
384	
385	
386	
387	
388	
389	
390	
391	
392	
393	
394	
395	
396	
397	
398	
399	
400	
401	
402	
403	
404	
405	
406	
407	
408	
409	
410	
411	
412	
413	
414	
415	
416	
417	
418	
419	
420	
421	
422	
423	
424	
425	
426	
427	
428	
429	
430	
431	
432	
433	
434	
435	
436	
437	
438	
439	
440	
441	
442	
443	
444	
445	
446	
447	
448	
449	
450	
451	
452	
453	
454	
455	
456	
457	
458	
459	
460	
461	
462	
463	
464	
465	
466	
467	
468	
469	
470	
471	
472	
473	
474	
475	
476	
477	
478	
479	
480	
481	
482	
483	
484	
485	
486	
487	
488	
489	
490	
491	
492	
493	
494	
495	

[illegible]

Db 76 ACCCC-TTTGGTCCCTCCCTTTGTATGGAGCTCTGTTTCATGTAATTCCTACTAT 19
 QY 364 TAAATCTGGCAACTGCA 380

Db 18 TAAATCTTGCAACTGCA 2

BI087886	LOCUS	DEFINITION	771 bp	mRNA	linear	EST 20-JUN-2001
BI087886	602852690F1	NIH_MGC_10 Homo sapiens		CDNA clone IMAGE:4993894	5',	mRNA sequence.

ACCESSION	BI087886
VERSION	BI087886.1
KEYWORDS	GI:14506216
SOURCE	EST.
	human.

ORGANISM Homo sapiens
Euryarchaeota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 771)
NIT-MGC <http://mc.manuscriptcentral.com/nit-mgc>
REFERENCE
AUTHORS

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@nsi-riemail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Distribution: WOT allows distribution information can be

clone distribution: no clone distribution information can be found through the I.M.A.G.E. Consortium/INL at: <http://image.lnl.gov>
 Plate: L1AM11015 row: d column: 23
 High quality sequence stop: 762.

```
FEATURES      Location/Qualifiers
      source    1..771
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:4993894"
              /clone_1pb="NIH_MGC_10"
```

```

/cell_line="WGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: PCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```

BASE COUNT 225 a 206 c 166 g 174 t

Query Match 68.9%; Score 345.2; DB 13; Length 771;
 Best Local Similarity 97.4%; Pred. No. 2.6e-101;
 Matches 372; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

Qy 1 GTCTGGCCAACTCCCAACAGCACTTAGGTCTTCTGT-TGAGATGGGGAGTGAAG 59
 Db 377 GTCTGGCCAACTCCCAACAGCACTTAGGTCTTCTGTGTGAGATGGGGAGTGAAG 436
 Qy 60 ACAGACTAGCTGGAATTTCTTAGCTGACTAGATCCCTTAGCTGAGTGGGAAGTGA 119
 Db 437 ACAGACTAGCTGGAATTTCTTAGCTGACTAGATCCCTTAGCTGAGTGGGAAGTGA 496
 Qy 120 CCACATCACTTTAAACAAGGGGCTTGCACTTAGCTCACTGACCAATCAGAGAGC 179
 Db 497 CCACATCACTTTAAACAAGGGGCTTGCACTTAGCTCACTGACCAATCAGAGAGC 556
 Qy 180 TCCTAAATGCTAAATTTAGCAAGAGAGTAAAGAAATGCAATCATCTATTGC 238
 Db 557 TCCTAAATGCTAAATTTAGCAAGAGAGTAAAGAAATGCAATCATCTATTGC 616
 Qy 239 CTGAGAGCAGCAGAGAGAGCAATGATCGGATATTAACCAAGTCTTCGAGCCGCA 298
 Db 617 CTGAGAGCAGCAGAGAGAGCAATGATCGGATATTAACCAAGTCTTCGAGCCGCA 676
 Qy 299 CGGCAACCCCTTTGGGTCCCTCTGTATGATGATGATGATGATGATGATGATGAT 358
 Db 677 CGGCAACCCCTTTGGGTCCCTCTGTATGATGATGATGATGATGATGATGATGAT 736
 Qy 359 TCTATTAATCTTGCACTGCA 380
 Db 737 TCTATTAATCTTGCACTGCA 758

RESULT 9
 AG102951 683 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-106G16.F, genomic survey sequence.
 DEFINITION AG102951
 ACCESSION AG102951.1 GI:16723468
 VERSION GSS.
 KEYWORDS Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male
 SOURCE BAC library clone:PTB-106G16.F.
 ORGANISM Pan troglodytes

REFERENCE
 AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Tokoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 JOURNAL 2 (bases 1 to 683)
 TITLE Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Tokoki,Y., Watanabe,H. and Sakaki,Y.
 JOURNAL Direct Submission
 JOURNAL Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan
 1-1-22 Suohiro-cho, Tsunuma-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimp@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the Red process and may have higher chance of
 clone tracking errors.
 PRIMERS
 Sequencing: -21M13

```

LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. .683
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-106G16.F"
/sex="male"
/cell_type="lymphoblast"
/clone_1lb="PTB Chimpanzee Male BAC library"

```

BASE COUNT 162 a 186 c 167 g 167 t

Query Match 68.8%; Score 343.8; DB 17; Length 683;
 Best Local Similarity 89.8%; Pred. No. 7.1e-101;
 Matches 403; Conservative 0; Mismatches 42; Indels 4; Gaps 3;

Qy 52 ACTGAGACAGAGCTAGCTGATTTCTTAGCTGACTAGCAATCCCTAGCTAGCTGG 111
 Db 108 ATTGAGAGAGAGCTAGCTGATTTCTTAGCTGACTAGCAATCCCTAGCTAGCTGG 167
 Qy 112 GAAGTGACCATTCACCTTTAAACAAGGGGCTTGCAACTTACTCAGACCTAGCAT 171
 Db 168 GAAGTGACCATTCACCTTTAAACAAGGGGCTTGCAACTTACTCAGACCTAGCAT 227
 Qy 172 CAGAGAGCTCACTAAATGCTAATTTAGCAAGAGAGAGTAAAGAAATAGCAATCAT 231
 Db 228 CAGAGAGCTCACTAAATGCTAATTTAGCAAGAGAGAGTAAAGAAATAGCAATCAT 286
 Qy 232 CTATTGCTGAGACAGCAGAGAGAGCAATGATCGGATATTAACCAAGTCTTGAG 291
 Db 287 CTATTGCTGAGACAGCAGAGAGAGCAATGATCGGATATTAACCAAGTCTTGAG 346
 Qy 292 CCGGCAACGGCAACCCCTTTGGGTCCCTCTGTATGATGATGATGATGATGATGAT 351
 Db 347 CCGGCAACGGCAACCCCTTTGGGTCCCTCTGTATGATGATGATGATGATGATGATGAT 404
 Qy 352 ATTTCACCTAATTAATCTTGCACTGACTCTGATGATGATGATGATGATGATGATGAT 411
 Db 405 ATTTCACCTAATTAATCTTGCACTGACTCTGATGATGATGATGATGATGATGATGAT 464
 Qy 412 CTGAGCTTTGCTGCGCATCCACCATGCTG-TTTCGCGCACCGCAGACCCGCGCTGA 470
 Db 465 CTGAGCTTTGCTGCGCATCCACCATGCTGTTTGTTCATATGACAGACCCGCGCTGA 524
 Qy 471 CTCCCATCCCTCTGATCATCAGGCTGT 499
 Db 525 CTTCATCCCTCCGATCCACAGAGTGT 553

RESULT 10
 N55091/c 385 bp mRNA linear EST 28-JAN-1997
 LOCUS IMAGE:245500 3', mRNA sequence.
 DEFINITION N55091
 ACCESSION N55091
 VERSION N55091.1 GI:1197970
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chippelli,B.,
 Chisoe,S., Dietrich,N., Dubuque,T., Favell,A., Gish,W., Hawkins
 'M., Hultman,M., Kucaba,T., Lacy,M., Le,N., Mardis,E., Moore
 'B., Morris,M., Parsons,J., Prange,C., Rife,J., Rohlfing,T.,
 Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevisan,E.,
 Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)

QY	304	ACCCCTTTGGGCTCCCTCCCTTTGATGGAGGCTCTGTTTCATGCTATTCACCTAT	363
Db	59	A-CCCCCTTTGGGCTCCCTCCCTTTGATGGAGGCTCTGTTTCATGCTATTCACCTAT	1
RESULT	12		
LOCUS	AM971553/c	522 bp	mRNA linear EST 01-JUN-2000
DEFINITION	EST383642	MAGE resequences, MAGI, Homo sapiens	CDNA, mRNA sequence.
ACCESSION	AM971553		
VERSION	AM971553.1	GI:8161399	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 522)		
AUTHORS	Hedde, P., Qi, R., Abernathy, K., Daarap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.		
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element CDNA microarray		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: John Quackenbush The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 3528 Fax: 301 838 0208 Email: john@igf.org Plate: 292		
FEATURES	Seq primer: Forward.		
source	Location/Qualifiers		
	1..522		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_lib="MAGE resequences, MAGI"		
	/note="Vector: pBluescriptSKm"		
BASE COUNT	125 a 115 c 134 g 143 t		
ORIGIN			
Query Match	66.3%; Score 332.4; DB 10; Length 522;		
Best Local Similarity	93.0%; Pred. No. 3.3e-97;		
Matches 348; Conservative	0; Mismatches 26; Indels 0; Gaps 0.		
QY	4	TCGGCCAACTCCCAACAGACATTAGTTTCCTCTTGATATGGGGAGCTGAGACAG	63
Db	374	TCAGCCAACTCACCACAAAGCACTGGTTTCCTCTTGATGAGAGGAGCTGAGACAG	315
QY	64	GACTAGCTGATTTCTTAGCTGACTAAGATTCCTAAGCCTTAGCTGGAGAGTGCAC	123
Db	314	GACTAGCTGATTTCTTAGCGGATTAAGATTCCTTAACCTTAGCTGGAGAGTGCAC	255
QY	124	ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAACCTGACCAATGAGAGCTCAC	183
Db	254	ATCCACCTTTAAACACGGGGCTTGCACTTAGCTCAACCTGACCAATGAGAGCTCAC	195
QY	184	TAAATGCTAATTAGGCAAGACAGAGGTAAAGAAATAGCCATCATATTGCTCGAG	243
Db	194	TAAATGCTAATTAGGCAAAACAGAGGTAAAGAAATAGCCATCATATTGCTCGAG	135
QY	244	AGCACGACGAGAGGACAAATGATCGGATATATAACCTAAGCTTGGAGCGGACGCA	303
Db	134	AGCACGATGGAGGACAGAGATTCGAATATATAACCAAGCATTCGAGCAGAACGGCA	75
QY	304	ACCCCTTTGGGCTCCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACCTAT	363
Db	74	ACCCCTTTGGGCTCCCTCCCTTTGATGGAGCTCTGTTTCATGCTATTCACCTAT	15
QY	364	TAAATCTGCACT 377	
Db	14	TAAATCTGCACT 1	

RESULT	13
T69704/c	
LOCUS	Y013A03.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:66988 3', mRNA sequence.
ACCESSION	T69704
VERSION	T69704.1 GI:680852
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 363) Hillier L., Clark L., Duboue T., Elliston K., Hawkins M., Holman M., Hulman M., Kucaba T., Le M., Lennon G., Marra M., Parsons J., Rifkin L., Rohlfing T., Soares M., Tan F., Trevaskis E., Waterston R., Williamson A., Wohldmann P. and Wilson R. The WashU-Merck EST Project Unpublished (1995)
AUTHORS	Contact: Wilison RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu
TITLE	Insert Size: 766
JOURNAL	High quality sequence stops: 341 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
COMMENT	Insert Length: 766 Std Error: 0.00 Seq primer: -21ml3 High quality sequence stop: 341. Location/Qualifiers 1 . 363 /organism="Homo sapiens" /db_xref="GDB:463693" /db_xref="taxon:9606" /clone="IMAGE:66988" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="VDH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pTV73D (pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGTGAAGATTAATTAAAGACTTTTCTTTTTTTTTCCTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo."
BASE COUNT	83 a 83 c 87 g 107 t 3 others
ORIGIN	
Query Match	65.7%; Score 329; DB 14; Length 363;
Best Local Similarity	97.3%; Pred. No. 3.6e-96;
Matches	354; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
Oy	17 CCACAGCATTTAGTTTCTCTGTGGAGTGCGGACCTGAGACAGCACTTAGCTGATT 76
Dd	363 CCAACAGCATTAGTTTCCGTGTGAATGGGGACGTAGAGCAGCATAGCTGATT 304
Oy	77 TCCTAGGCTGACTAGATTCCTTAAGCCTAGCTGGGAAGGTGACCAACATCCACTTTAAA 136
Dd	303 TCTTAGGCTGATAAGATCCCTAGGCTAGTGGGAAGGTGACCACATCCACTTTAAA 244
Oy	137 CACGGGGCTTGCAATTAGTCTCACACCTGCACCAATCAGAAGCTCCTAAATGTCTAAT 196
Dd	243 CACGGGGCTTGCAATTAGTCTCACACCTGCACCAATCAGAAGCTCCTAAATGTCTAAT 184
Oy	197 AGGCAAGACAGAGGTAAAGAAATAGCCAATCATCTCTTGGCTGAGAGCACAGCAGAG 256

Db 183 AGCAAAAGACAGAGGTAAAGAAATAGCCATATCTATTGCTGAGACAGACAGAG 124

QY 257 GGAATATGATCGGATATTAACCCAGTCTTCAGCGGCAACGCCATCCCTTTGGGT 316

Db 123 GGAATATGATCGGATATTAACCCAGTCTTCAGCGGCAACGCCATCCCTTTGGGT 66

QY 317 CCCCTCCCTTTGATGAGAGCTCTGTTTTCATGCTATTTCACTCTATTAATCTTGAAC 376

Db 65 CCCCTCCCTTTGATGAGAGCTCTGTTTTCATGCTATTTCACTCTATTAATCTTGAAC 6

QY 377 TGCA 380

Db 5 NACA 2

RESULT 14

AA860368/c 342 bp mRNA linear EST 31-DEC-1998

LOCUS aJ59c05.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394600

DEFINITION 3' similar to contains PRR7.c1 PRR5 repetitive element ;, mRNA sequence.

ACCESSION AA860368 GI:2954363

VERSION AA860368.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 342)

TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

COMMENT Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

EMAIL: cgaaps-remail.nih.gov

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bernaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

CDNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www.bio.lnl.gov/dbtrp/image/image.html

Insert Length: 1305 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 319.

Location/Qualifiers

1..342

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1394600"

/clone_1lb="Soares_testis_NHT"

/sex="male"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bernaldo.

BASE COUNT 75 a 78 c 82 g 107 t

ORIGIN

Query Match 65.2%; Score 326.6; DB 9; Length 342;

Best Local Similarity 98.8%; Pred. No. 2,2e-95;

Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 48 GGGGACTGAGACAGAGCTAGTGAATTTCTTGGCTGACTAAGAAATCCCTAAGCTTAA 107

Db 342 GGGGACTGAGACAGAGCTAGTGAATTTCTTGGCTGACTAAGAAATCCCTAAGCTTAA 283

QY 108 CTGGGAAGGTGACACATCCACCTTTTAAACAGCGGGCTTGCATTAAGCTACACCTGAC 167

Db 282 CTGGGAAGGTGACACATCCACCTTTTAAACAGCGGGCTTGCATTAAGCTACACCTGAC 223

QY 168 CAATCAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCCAA 227

Db 222 CAATCAGAGCTCACTAAATGCTAATTTAGGCAAGACAGAGGTAAAGAAATAGCCAA 163

QY 228 TCATCTATTCCTGAGAGCAGAGAGGAGCAATGATCGGATATTAACCCAGCTT 287

Db 162 TCATCTATTCCTGAGAGCAGAGAGGAGCAATGATCGGATATTAACCCAGCTT 103

QY 288 CGAGCCGCAACGCAACCCCTTTGGGTCCCTCCCTTTGATGAGAGCTCTGTTTCA 347

Db 102 CGAGCCGCAACGCAACCCCTTTGGGTCCCTCCCTTTGATGAGAGCTCTGTTTCA 43

QY 348 TGCTATTTCACTCTAATTAATCTTGAACCTGCA 380

Db 42 TGCTATTTCACTCTAATTAATCTTGAACCTGCA 10

RESULT 15

R27412/c 424 bp mRNA linear EST 24-APR-1995

LOCUS yH46d11.s1 Soares_placenta_Nb2HP Homo sapiens cDNA clone

DEFINITION IMAGE:132789 3', mRNA sequence.

ACCESSION R27412 GI:783547

VERSION R27412.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 424)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maira, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilson RK

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

Insert Size: 695

High quality sequence stops: 367

Source: IMAGE Consortium, LINL

This clone is available royalty-free through LINL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 695 Std Error: 0.00

Seq primer: Promega -21m13

High quality sequence stop: 367.

Location/Qualifiers

1..424

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:132789"

/clone_1lb="Soares_placenta_Nb2HP"

/sex="Female"

/dev_stage="Placenta obtained at birth (full term)"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGGCGCGCAGAGAAATTTTCTTTTCTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization.

BASE COUNT 92 a 96 c 115 g 121 t constructed by Bento Soares and M.Fatima Bonaldo. "
ORIGIN

Query Match 65.1%; Score 326.4; DB 14; Length 424;
Best Local Similarity 93.9%; Pred. No. 2.8e-95;
Matches 351; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

QY 4 TCGGCGAACCTCCCGCAACAGACACTAGGTTTCCCTGTTGAGATGGGGACTGAGACAG 63
DB 373 TCGATCAACCTCCCGCAACAGACACTAGGTTTCCCTGTTGAGATGGGGACTGAGACAG 314
QY 64 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCTAGCTGGGAAGTGACAC 123
DB 313 GACTAGCTGATTTCTTAGGCTGACTAAGATCCCTAAGCTAGCTGGGAAGTGACTGC 254
QY 124 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCAGACCTGACCAATCAGAGAGCTCAG 183
DB 253 ATCCACCTTTAAACAGGGGCTTGCACTTAGCTCAGACCTGACCAATC--AGAGCTCAG 196
QY 184 TAAATGCTAATTAAGCAAGAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTGAG 243
DB 195 TAAATGCTAATTAAGCAAGAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTGAG 136
QY 244 AGCAGAGCAGAGAGGAGATGATCGGATATTAACCTCAAGTCTTGAGCCGCGCA 303
DB 135 AGCAGAGCAGAGAGGAGATGATCGGATATTAACCTCAAGTCTTGAGCCGCGCA 76
QY 304 ACCGCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCATGCTATTTCACCTAT 363
DB 75 AACCCCTTTGGGTCCCTCCCTTTGTATGGAGCTCTGTTTCATGCTATTTCACCTAT 16
QY 364 TAAATCTGCAACT 377
DB 15 TAAATCTGCAACT 2

Search completed: April 17, 2003, 07:35:33
Job time : 1270.84 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:07:23 ; Search time 1703.47 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Sequence: 1 caagatccgcagatcataca.....tcagtgcagacacatccag 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
GenEmbl:
1:  gb_da:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_ph:*
7:  gb_pat:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	6 AX007980	AX007980 Sequence
2	501	100.0	56093	6 AX329572	AX329572 Sequence
3	501	100.0	56093	9 HSAAC000064	AC000064 Human BAC
4	501	100.0	149194	9 AC007566	AC007566 Homo sapi
5	474	94.6	2938	6 AX000957	AX000957 Sequence
6	474	94.6	2938	6 AX027471	AX027471 Sequence
7	474	94.6	2938	6 AF072499	AF072499 Homo sapi
8	442	88.2	7582	6 AX000966	AX000966 Sequence
9	442	88.2	7582	6 AX027480	AX027480 Sequence
10	435	86.8	40205	9 AF045450	AF045450 Homo sapi
11	435	86.8	142742	9 AF121782	AF121782 Homo sapi
12	435	86.8	340000	9 HS21C080	AL163280 Homo sapi
13	428.6	85.5	8339	9 AL162912	AL162912 Human DNA
14	428.4	85.5	3372	6 AX000964	AX000964 Sequence
15	428.4	85.5	3372	6 AX027478	AX027478 Sequence
16	428.4	85.5	3372	6 AF072502	AF072502 Homo sapi
17	425.4	84.9	146545	2 AC104009	AC104009 Homo sapi
18	423.8	84.6	143063	9 AC040936	AC040936 Homo sapi
19	423.8	84.6	163166	9 AL356632	AL356632 Human DNA
20	423.8	84.6	176773	2 AC009867	AC009867 Homo sapi
21	422.2	84.3	2304	14 AF009668	AF009668 Multiple
22	422.2	84.3	108232	9 AP000654	AP000654 Homo sapi
23	422.2	84.3	156938	9 AL159163	AL159163 Human DNA
24	422.2	84.3	184523	2 AP002890	AP002890 Homo sapi
25	421.2	84.1	139744	9 HS83513	AL133513 Human DNA
26	420.6	84.0	132981	9 HSR2J11	283850 Human DNA
27	420.6	84.0	161771	9 CENS1DRX	AL121579 Human chr
28	420.6	84.0	178152	2 AC108746	AC108746 Homo sapi
29	420.6	84.0	191426	2 AC026086	AC026086 Homo sapi
30	420.6	84.0	194545	2 AC087482	AC087482 Homo sapi
31	420.6	84.0	194718	2 AC026977	AC026977 Homo sapi
32	420.6	84.0	220807	9 AC093334	AC093334 Homo sapi
33	420.6	84.0	251124	9 HUA600660	HE000660 Homo sapi
34	419.6	83.8	1158	6 A46012	A46012 Sequence 1
35	419.6	83.8	1158	6 A57048	A57048 Sequence 1
36	419.6	83.8	1158	6 A60113	A60113 Sequence 1
37	419.6	83.8	1158	6 A79463	A79463 Sequence 1
38	419.6	83.8	1158	6 A80197	A80197 Sequence 1
39	419.6	83.8	1158	6 A80306	A80306 Sequence 1
40	419.6	83.8	1158	6 AR035349	AR035349 Sequence
41	419.6	83.8	1158	6 AR036190	AR036190 Sequence
42	419.6	83.8	1158	6 AR036710	AR036710 Sequence
43	419.6	83.8	1158	6 AR076972	AR076972 Sequence
44	419.6	83.8	1158	6 AR094693	AR094693 Sequence
45	419.6	83.8	1158	6 AR183928	AR183928 Sequence

ALIGNMENTS

RESULT 1
AX007980 10499 bp DNA linear PAT 06-SEP-2000
LOCUS Sequence 3 from Patent WO9967395.
DEFINITION AX007980
ACCESSION
VERSION AX007980.1 GI:9995677
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
TITLE Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

JOURNAL

Patent: WO 9667395-A 3 29-DEC-1999;
INST NAT SANTE RECH MED (FR); FERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)

FEATURES

Location/Qualifiers

1..10499
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 3048 a 2676 c 2280 g 2195 t
ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 10499;

Best Local Similarity 100.0%; Pred. No. 2.3e-138; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatch:hes 0; Indels 0; Gaps 0;

```
Oy 1 CAAGATCTCAGGATTTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACCTTACCT 60
    |||||
Db 5000 CAGATCTCAGGATTTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACCTTACCTT 5059
Oy 61 TATACCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
    |||||
Db 5060 TATACCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
Oy 121 GATGCTCTTCTGATCCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 180
    |||||
Db 5120 GATGCTCTTCTGATCCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 5179
Oy 181 ACTTCAAAACCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 240
    |||||
Db 5180 ACTTCAAAACCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 5239
Oy 241 CCCCATCTATTGTCAGGAGGATTAAGCCCAAGACTTGAAGCAATCTCATACCTGAGCACT 300
    |||||
Db 5240 CCCCATCTATTGTCAGGAGGATTAAGCCCAAGACTTGAAGCAATCTCATACCTGAGCACT 5299
Oy 301 TGTCCTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    |||||
Db 5300 TGTCCTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5359
Oy 361 AGCCACCCAGAGGCTCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
    |||||
Db 5360 AGCCACCCAGAGGCTCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5419
Oy 421 GGCTCAACTCTGCTGACAGCAGGTTACTTGGGCTTAAATTTTCCAAAGACCAAGGGCC 480
    |||||
Db 5420 GGCTCAACTCTGCTGACAGCAGGTTACTTGGGCTTAAATTTTCCAAAGACCAAGGGCC 5479
Oy 481 CTCAGTGAGGAACATCCAG 501
    |||||
Db 5480 CTCAGTGAGGAACATCCAG 5500
```

RESULT 2

AX329572 56093 bp DNA linear PAT 09-JAN-2002

LOCUS AX329572 Sequence 81 from Patent WO0194629.

ACCESSION AX329572

VERSION AX329572.1 GI:18102550

KEYWORDS

SOURCE

ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Hortigian, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 81 13-DEC-2001;

TITLE

JOURNAL

Avalon Pharmaceuticals (US)

FEATURES

SOURCE

1..56093
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t
ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 56093;

Best Local Similarity 100.0%; Pred. No. 2.3e-138; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatch:hes 0; Indels 0; Gaps 0;

```
Oy 1 CAAGATCTCAGGATTTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACCTTACCT 60
    |||||
Db 33000 CAGATCTCAGGATTTATCATAGAGCTGTGTTCTCTATACCGAGCTGTACCTTACCTT 33059
Oy 61 TATACCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
    |||||
Db 33060 TATACCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 33119
Oy 121 GATGCTCTTCTGATCCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 180
    |||||
Db 33120 GATGCTCTTCTGATCCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 33179
Oy 181 ACTTCAAAACCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 240
    |||||
Db 33180 ACTTCAAAACCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 33239
Oy 241 CCCCATCTATTGTCAGGAGGATTAAGCCCAAGACTTGAAGCAATCTCATACCTGAGCACT 300
    |||||
Db 33240 CCCCATCTATTGTCAGGAGGATTAAGCCCAAGACTTGAAGCAATCTCATACCTGAGCACT 33299
Oy 301 TGTCCTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
    |||||
Db 33300 TGTCCTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 33359
Oy 361 AGCCACCCAGAGGCTCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
    |||||
Db 33360 AGCCACCCAGAGGCTCTTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 33419
Oy 421 GGCTCAACTCTGCTGACAGCAGGTTACTTGGGCTTAAATTTTCCAAAGACCAAGGGCC 480
    |||||
Db 33420 GGCTCAACTCTGCTGACAGCAGGTTACTTGGGCTTAAATTTTCCAAAGACCAAGGGCC 33479
Oy 481 CTCAGTGAGGAACATCCAG 501
    |||||
Db 33480 CTCAGTGAGGAACATCCAG 33500
```

RESULT 3

HSAC000064 56093 bp DNA linear PRI 13-NOV-1996

LOCUS HSAC000064 Human BAC clone RG083M05 from 7q21-7q22, complete sequence.

DEFINITION AC000064

ACCESSION AC000064.1 GI:1669369

KEYWORDS

SOURCE

ORGANISM

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 56093)
The sequence of H. sapiens BAC clone RG083M05
Paulley, A.

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (13-NOV-1996)
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63108, USA

e-mail: saplens@wustl.edu

NOTICE: This sequence may not represent the entire insert of this
section. It may be shorter because we only sequence overlapping
sections once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBELO

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H_RG083M05; actual end is at 56093 of H_RG083M05

This clone contains STS SMS1725.

Location/Qualifiers

1. 56093

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone_id="H_RG083M05"

/clone_1id="CITB-978SK-B"

/repeat_region (838..1131)

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(<1360..1503,4181..4370,4587..4774,6422..6556,

9483..9547,11631..11773,11864..12021,13131..13296,

14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="Atpase; strong similarity to peroxisome

biosynthesis protein PAB1 (PID:g1172019); coded for by

human cDNA C04279 (NID:g1467550)"

/codon_start=1

/protein_id="PAB46346.1"

/db_xref="GI:1669371"

/translation="KRLEINQKTELEVAFAVMNOPSVLDDLDLIAGLPAVEREH

SPABORCEILNVIKKIDCDINKPTDILQVAKETGFAVDPTVADRAIHRRL

SRSGISTREKLVTTIDPQKALRGFPAPASIRSNLHKPRGLGDKIGGHEHQIILMD

TIQPAKVCIAKKEKYPBELFANLPRIQRTGILGYPETGTLIAGVIAESRNFISV

KGBELSKYIGABEQAVRDIPIAQAQPCILFDEDESIAPRGHQNTGTVRVNQ

LIQLDVEGLQGVYLAATSRDLIDPALLRGRLDCYCPDPQDITSYLESKTQ

OMHSFVSRLEILINVLDSLPLADVDLQHVASVDSFGADLKALLVAOLEALHG

MLSKMSKELPDESKFMVRLYFGSSYVESLNGTSDLSGLSA PSSMTDLPVPI

GKQQLSQPVATFASQEQCELTQEBRDLPADISIKRRTYSGSGEDSMNQPGPI

KIRLAISQSHLMTALGHTRPISDDMKNAEL"

/complement (4948..5130)

/rpt_family="ALU"

/complement (6581..7133)

/rpt_family="U1"

/complement (7767..8037)

/rpt_family="ALU"

/complement (8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:g1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:g1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N2627 (NID:g1130501), bases

276-343"

repeat_region

complement (12612..12907)

/rpt_family="ALU"

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

143-266"

13794..13877

/rpt_family="ALU"

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:g17434), bases

30-58"

13907..14104

/rpt_family="ALU"

complement (14110..14137)

/rpt_family="U1"

complement (15618..15907)

/rpt_family="ALU"

17227..17522

/rpt_family="ALU"

16667..19235

/note="match to human fetal brain 5' EST D61494

(NID:g970409), bases 1-255, and to human 3' EST R07476

(NID:g973939)"

19550..19670

/rpt_family="ALU"

21507..37303

/note="similarity to various SS-RNA virus polyproteins;

pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:g942257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

complement (38938..39224)

/rpt_family="ALU"

39225..39707

/note="match to multiple human ESTs, see N30113

(NID:g1148633)"

39800..40085

/rpt_family="ALU"

complement (40247..40538)

/rpt_family="ALU"

complement (40632..40924)

/rpt_family="ALU"

complement (42283..42891)

/rpt_family="ALU"

complement (45474..45613)

/rpt_family="ALU"

complement (45614..45737)

/note="match to human 3' EST H48898 (NID:g988738), bases

139-333"

complement (46107..47026)

/note="match to multiple human ESTs, see N81064

(NID:g1243765), H48897 (NID:g988737), and M78831

(NID:g273146)"

complement (47027..47318)

/rpt_family="ALU"

complement (47365..47782)

/note="match to multiple human ESTs, see W37495

(NID:g1319089)"

47898..48115

/note="match to human 5' EST H62306 (NID:g1015138), bases

93-368"

complement (48116..48405)

/rpt_family="ALU"

complement (48406..48584)

/note="match to human 3' EST N29952 (NID:g1148472), bases

230-455, and 5' EST R12730 (NID:g765806)"

complement (48787..49405)

/rpt_family="ALU"

complement (49406..49534)

/note="match to human 3' EST R65794 (NID:g838432), bases

309-440"

```

repeat_region      complement(49638..49672)
                    /rpt_family="ALU"
misc_feature        /note="match to human 3' EST N29952 (NTD:g1148472) and 5'
                    EST N29938 (NTD:g1148458), sequences are from opposite
                    ends of the same clone"
gene                /gene="WUGSC:H.RG083M05.2"
                    /complement(49698..51306)
                    /note="WUGSC:H.RG083M05.2"
CDS                 /complement(49698..49888,51575..51806)
                    /gene="WUGSC:H.RG083M05.2"
                    /note="coded for by human cDNA M37389 (NTD:g119205),
                    R65891 (NTD:g838529), R65794 (NTD:g838432) and R65794
                    (NTD:g838432)"
                    /codon_start=1
                    /protein_id="AAB46345.1"
                    /db_xref="GI:1669370"
                    /translation="MFTYRQCGIIFPCPGVYVQIGDVSVIDEQGRPYAIRF
                    IDQYCEKSAALTWLITPLSSPRQPDASTYIGPEEDLPKMYLEFVCHAPSEYRK
                    SRSSPFPVTPRPEKGYIWHV3PTPALTIKESVANHL"
                    complement(51576..51738)
exon                 /gene="WUGSC:H.RG083M05.2"
                    /note="Grail Prediction, score = 86"
                    /evidence=not_experimental
repeat_region       complement(52052..52339)
                    /rpt_family="L1"
                    55557..55843
misc_feature         /note="match to human HST M79192 (NTD:g273505) base 2-289"
                    100.0%; Score 501, DB 9, Length 56093;
Query Match         Best Local Similarity 100.0%; Pred. No. 2,3e-138;
Matches 501, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

```

1 CAAGTCTCAGATTATTCATGAGCTGTTGCTCTATACCCAGCTGTACTGACCT 60
 Db 33000 CAGATCTCAGATTATTCATGAGCTGTTGCTCTATACCCAGCTGTACTGACCT 33059
 Qy 61 TATACCTGCTTTCCCAATACACAGAGAGAGAGGTTTACAGTCTCTGACCTTCAG 120
 Db 33060 TATACCTGCTTTCCCAATACACAGAGAGAGAGGTTTACAGTCTCTGACCTTCAG 33119
 Qy 121 GATGCTCTCTTTCGATCCTCTGTACATCTGACTTTCATTTCTTTGTTGCTTTGAAGT 180
 Db 33120 GATGCTCTCTTTCGATCCTCTGTACATCTGACTTTCATTTCTTTGTTGCTTTGAAGT 33179
 Qy 181 ACTTCAAAACCAACATCTCACTGACCTGATATTTTACCCCAAGGTTTACAGGATAGT 240
 Db 33180 ACTTCAAAACCAACATCTCACTGACCTGATATTTTACCCCAAGGTTTACAGGATAGT 33239
 Qy 241 CCCATCTATTTGGCCAGGATTAAGCCCAAGACTTACGCCAATCTCTCATCCTGACACT 300
 Db 33240 CCCATCTATTTGGCCAGGATTAAGCCCAAGACTTACGCCAATCTCTCATCCTGACACT 33299
 Qy 301 TGTCTTGGTGAAGTGATGATTTACTTTGGCCGTCATTAGAGAACTTTGTGCATCA 360
 Db 33300 TGTCTTGGTGAAGTGATGATTTACTTTGGCCGTCATTAGAGAACTTTGTGCATCA 33359
 Qy 361 AGCCACCAAGAGGCTCTTCAATTTCTGCTACCTCTGTGCTACATGTTTCCAAACCAA 420
 Db 33360 AGCCACCAAGAGGCTCTTCAATTTCTGCTACCTCTGTGCTACATGTTTCCAAACCAA 33419
 Qy 421 GGCTCAACTGTGCTCAACAGAGTTACTTGGGCTAAATATTCAGAGGACCAAGGCGC 480
 Db 33420 GGCTCAACTGTGCTCAACAGAGTTACTTGGGCTAAATATTCAGAGGACCAAGGCGC 33479
 Qy 481 CTCAGTGAAGAACATCTCAG 501
 Db 33480 CTCAGTGAAGAACATCTCAG 33500

RESULT 4
 AC007566/c AC007566 149194 bp DNA linear PRI 01-MAR-2002
 LOCUS Homo sapiens BAC clone CTB-1065 frcm 7q21-7q22, complete sequence.
 DEFINITION

```

ACCESSION          AC007566
VERSION            AC007566.2
KEYWORDS           GI:11181861
SOURCE             Homo sapiens.
ORGANISM           Homo sapiens
REFERENCE           Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
                   1 (bases 1 to 149194)
                   Sultoni,J.E. and Waterston,R.
                   Genome Res. 8 (11), 1097-1108 (1998)
                   2 (bases 1 to 149194)
                   Du,Z.
                   The sequence of Homo sapiens BAC clone CTB-1065
                   Unpublished (2001)
                   3 (bases 1 to 149194)
                   Waterston,R.H.
                   Direct Submission
                   Submitted (15-MAY-1999) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
                   4 (bases 1 to 149194)
                   Waterston,R.
                   Direct Submission
                   Submitted (02-OCT-2000) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
                   5 (bases 1 to 149194)
                   Waterston,R.H.
                   Direct Submission
                   Submitted (16-NOV-2000) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
                   6 (bases 1 to 149194)
                   Waterston,R.H.
                   Direct Submission
                   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
                   7 (bases 1 to 149194)
                   Waterston,R.H.
                   Direct Submission
                   Submitted (06-FEB-2002) Genome Sequencing Center, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
                   On Nov 16, 2000 this sequence version replaced gi:4835815.
                   ----- Genome Center
                   Center: Washington University Genome Sequencing Center
                   Center code: WUGSC
                   Web site: http://genome.wustl.edu/gsc
                   Contact: sapiens@wustl.wustl.edu
                   ----- Summary Statistics
                   Center project name: H_RG010G05

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NGR1 Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

```
1.149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_1lb="CTB-978SK-B"
1.1654
/rpt_family="L1"
misc_feature
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST B1160365 (NID:g14620366)"
misc_feature
2248..2374
/note="match to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
2696..3066
/rpt_family="MALR"
3108..3392
/rpt_family="Alu"
3540..3628
/note="match to EST AA425526 (NID:g12106267) zw48b03.r1"
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
3542..3628
/note="match to EST BG766882 (NID:g14077535)"
```

```
misc_feature 3542..3628
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 3542..3609
/note="match to EST AW579261 (NID:g7254310)"
misc_feature 3562..3628
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3598..3628
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 3717..3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature 3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature 3719..3785
/note="match to EST AA425526 (NID:g12106267) zw48b03.r1"
misc_feature 3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 3719..3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 3719..3785
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 3719..3785
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 3719..3785
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature 3860..3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region 4730..4827
/rpt_family="L2"
4834..4925
/rpt_family="L1"
misc_feature 5715..5860
/note="match to EST AA425526 (NID:g12106267) zw48b03.r1"
misc_feature 5715..5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature 5715..5860
/note="match to EST BE272564 (NID:g9146913)"
misc_feature 5715..5860
/note="match to EST BE299708 (NID:g9183456)"
misc_feature 5715..5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature 5715..5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature 5715..5860
/note="match to EST B1160365 (NID:g14620366)"
misc_feature 5715..5860
/note="similar to Mus musculus EST BE994936
```

Query Match 100.0%; Score 501; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 2.2e-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CAAGATCTCAGATTAATCATGAGGCTGTTCTTATATAGCCAGTGTACCTAGCCCT 60
DB 89004 CAAGATCTCAGATTAATCAATGAGGCTGTTCTTATATAGCCAGTGTACCTAGCCCT 88945
QY 61 TATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCCGAGCTTCAG 120
DB 88944 TATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCCGAGCTTCAG 88985
QY 121 GATGCTTCTTCTGATCCTGTATACCTGACTCTCAATTTCTTGTGCTTTGAAGAT 180
DB 88884 GATGCTTCTTCTGATCCTGTATACCTGACTCTCAATTTCTTGTGCTTTGAAGAT 88825
QY 181 ACTTCAACCAACATTCATCACTGATCTATTTTACCCCAAGGTTCCAGGATAGT 240
DB 88824 ACTTCAACCAACATTCATCACTGATCTATTTTACCCCAAGGTTCCAGGATAGT 88765
```

241 |CCCCATCTATTGGGCGAGCATTTAGCCCAAGAC.TGAGCCCAATCCTCATACCTGGACACT| 300
|CCCCATCTATTGGGCGAGCATTTAGCCCAAGAC.TGAGCCCAATCCTCATACCTGGACACT| 88705
|CCCCATCTATTGGGCGAGCATTTAGCCCAAGAC.TGAGCCCAATCCTCATACCTGGACACT| 88705
301 |TGTCCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATCA| 360
|TGTCCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATCA| 88704
|TGTCCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATCA| 88645
361 |AGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCAAA| 420
|AGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCAAA| 88644
|AGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCAAA| 88585
421 |GGCTCAACTGTCTCAGACAGAGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGCC| 480
|GGCTCAACTGTCTCAGACAGAGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGCC| 88584
|GGCTCAACTGTCTCAGACAGAGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGCC| 88525
481 |CTCAGTGAGGAACACATCCAG 501
|CTCAGTGAGGAACACATCCAG 501
|CTCAGTGAGGAACACATCCAG 501
88524 |CTCAGTGAGGAACACATCCAG 88504

RESULT 5
AX000957 2938 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 2 from Patent WO9902696
DEFINITION AX000957
ACCESSION AX000957
VERSION AX000957.1 GI:7241199
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
unclassified.
1 (bases 1 to 2938)
REFERENCE
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 2 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source location/Qualifiers
1. 2938
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

1 |CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTAGCCC| 59
|CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTAGCCC| 2079
|CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTAGCCC| 2138
60 |TTATACCTGCTTTCCCAATACAGAGAGACAG;GTGGTTAAGATCTGAGACCTTCA| 119
|TTATACCTGCTTTCCCAATACAGAGAGACAG;GTGGTTAAGATCTGAGACCTTCA| 2139
|TTATACCTGCTTTCCCAATACAGAGAGACAG;GTGGTTAAGATCTGAGACCTTCA| 2198
120 |GGATGCTCTTCTTGATGATCCCTGTACATCTGACTCTCAATTTCTTTGGCTTTGAAGA| 179
|GGATGCTCTTCTTGATGATCCCTGTACATCTGACTCTCAATTTCTTTGGCTTTGAAGA| 2199
|GGATGCTCTTCTTGATGATCCCTGTACATCTGACTCTCAATTTCTTTGGCTTTGAAGA| 2258
180 |TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 239
|TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 2259
|TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 2318
239 |TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 2318
240 |TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 299
|TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 2319
|TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 2378
300 |TTGTCTCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATC| 359
|TTGTCTCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATC| 2379
|TTGTCTCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATC| 2438
360 |AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 419
|AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 2379
|AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 2438

2439 |AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 2498
420 |AGGCTCAACTGTCTCAGACAGGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGC| 479
2499 |AGGCTCAACTGTCTCAGACAGGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGC| 2558
480 |CCTCAGTGAGGAACACATCCAG 501
|CCTCAGTGAGGAACACATCCAG 501
|CCTCAGTGAGGAACACATCCAG 501
2559 |CCTCAGTGAGGAACACATCCAG 2580

RESULT 6
AX027471 2938 bp DNA linear PAT 16-SEP-2000
LOCUS AX027471
DEFINITION Sequence 21 from Patent FR2788784.
ACCESSION AX027471
VERSION AX027471.1 GI:10188435
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Homo sapiens
Bakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Mallet, F., Voisard, C. and Paranhos, B.G.
JOURNAL Patent: FR 2788784-A 21 28-JUL-2000;
BIO MERIEUX (FR)
FEATURES
source location/Qualifiers
1. 2938
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

1 |CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTAGCCC| 59
|CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTAGCCC| 2079
|CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGACTAGCCC| 2138
60 |TTATACCTGCTTTCCCAATACAGAGAGACAG;GTGGTTAAGATCTGAGACCTTCA| 119
|TTATACCTGCTTTCCCAATACAGAGAGACAG;GTGGTTAAGATCTGAGACCTTCA| 2139
|TTATACCTGCTTTCCCAATACAGAGAGACAG;GTGGTTAAGATCTGAGACCTTCA| 2198
120 |GGATGCTCTTCTTGATGATCCCTGTACATCTGACTCTCAATTTCTTTGGCTTTGAAGA| 179
|GGATGCTCTTCTTGATGATCCCTGTACATCTGACTCTCAATTTCTTTGGCTTTGAAGA| 2199
|GGATGCTCTTCTTGATGATCCCTGTACATCTGACTCTCAATTTCTTTGGCTTTGAAGA| 2258
180 |TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 239
|TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 2259
|TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 2318
239 |TACTTCAAAACCAACATCTCACTGACTGATATTTTACCCCAAGGGTTAGGGGATAG| 2318
240 |TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 299
|TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 2259
|TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 2378
239 |TCCCATCTATTGGCGAGCATTTAGCCCAAGACCTTACCCATCTCATACCTGGACAC| 2378
300 |TTGTCTCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATC| 359
|TTGTCTCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATC| 2379
|TTGTCTCTTGGTAGGTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTGCGATC| 2438
360 |AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 419
|AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 2439
|AAGCCACCCCAAGGCTCTTCATTTTCTGCTACCTGTGGCTACATGTTTCCAAACCA| 2498
420 |AGGCTCAACTGTCTCAGACAGGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGC| 479
2499 |AGGCTCAACTGTCTCAGACAGGTTACTTAGGGCTAAATTAATCCAAAGCACAGGCGC| 2558
480 |CCTCAGTGAGGAACACATCCAG 501
|CCTCAGTGAGGAACACATCCAG 501
|CCTCAGTGAGGAACACATCCAG 501
2559 |CCTCAGTGAGGAACACATCCAG 2580

RESULT 7
AF072499 2938 bp mRNA linear PRI 10-FEB-1999
LOCUS Homo sapiens endogenous retrovirus W sequence.
DEFINITION AF072499
ACCESSION AF072499.1 GI:4262283
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2938)
AUTHORS Blond, J. L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H.,
Mandrard, B. and Mallet, F.
TITLE Molecular characterization and placental expression of HERV-W, a
new human endogenous retrovirus family
JOURNAL J. Virol. 73 (2), 1175-1185 (1999)
MEDLINE 99099005
PubMed 9882319
REFERENCE 2 (bases 1 to 2938)
AUTHORS Blond, J. L., Beseme, F. and Mallet, F.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1998) UM103 CNRS-biomerieux, ENS Lyon, 46 allée
d'Italie, Lyon, Cedex 07 69364, France
FEATURES
source location/Qualifiers
1..2938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Cl.6A1"
/cissue_type="placenta"
/clone_lib="Clontech 5'-stretch plus library, Cat number
HL0147"
misc_feature 1..2938
/note="corresponds to gag and pol"
repeat_region 1..2938
/rpt_family="HERV-W"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 9; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAAGATCTCAGGATTTAAT-GAGGCTGTGTTCTCTATAGCCAGTGAACCTGACC 59
DB 2079 CAAGATCTCAGGATTTAATGAGGCGGTGTCTTTATACCCAGCTGTAACCTGACC 2138
QY 60 TTATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCTGAGACCTTCA 119
DB 2139 TTATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCTGAGACCTTCA 2138
QY 120 GGATGCTTCTTCTGATCCCTGTACATCTGACTGCAATCTTGTGCTTGAAGA 179
DB 2199 GGATGCTTCTTCTGATCCCTGTACATCTGACTGCAATCTTGTGCTTGAAGA 2258
QY 180 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCAGGGTTCAAGGATAG 239
DB 2259 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCAGGGTTCAAGGATAG 2318
QY 240 TCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAGCCAACTCATACCTGACAC 299
DB 2319 CCCCACATATTTGGCCAGGATTAAGCCCAAGACTTGAGCCAACTCATACCTGACAC 2378
QY 300 TTGTCTCTCGGTAGGTGATGATTTTCTTTGGCCGCCCATTCAGAAAACCTTGTGCCATC 359
DB 2379 TTGTCTCTCGGTAGGTGATGATTTTCTTTGGCCGCCCATTCAGAAAACCTTGTGCCATC 2438
QY 360 AAGGCAACCAAGCGCTTTCATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 419
DB 2439 AAGGCAACCAAGCGCTTTCATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 2498

QY 420 AGGCTCACTCTGCTCAGAGAGGTTACTTAGGGCTAAATTTATCCAAAGCACAGGGC 479
DB 2499 AGGCTCACTCTGCTCAGAGAGGTTACTTAGGGCTAAATTTATCCAAAGCACAGGGC 2558
QY 480 CCTCAGTGAGGAACATCCAG 501
DB 2559 CCTCAGTGAGGAACATCCAG 2580

RESULT 8
AX000966
LOCUS AX000966 7582 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 11 from Patent WO9902696.
ACCESSION AX000966
VERSION AX000966.1 GI:7241208
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 7582)
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 11 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source location/Qualifiers
1..7582
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others
ORIGIN

Query Match 88.2%; Score 442; DB 6; Length 7582;
Best Local Similarity 91.3%; Pred. No. 1e-120;
Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;

QY 1 CAAGATCTCAGGATTTAAT-GAGGCTGTGTTCTCTATAGCCAGCTGTAACCTGACC 59
DB 2678 CAAGATCTCAGGATTTAATGAGGCGGTGTCTTTATACCCAGCTGTAACCTGACC 2737
QY 60 TTATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCTGAGACCTTCA 119
DB 2738 TTATATCTGCTTCCCAATATACAGAGAGAGAGGTTTACAGTCTGAGACCTTCA 2797
QY 120 GGATGCTTCTTCTGATCCCTGTACATCTGACTGCAATCTTGTGCTTGAAGA 179
DB 2798 GGATGCTTCTTCTGATCCCTGTACATCTGACTGCAATCTTGTGCTTGAAGA 2857
QY 180 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCAGGGTTCAAGGATAG 239
DB 2858 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCAGGGTTCAAGGATAG 2917
QY 240 TCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAGCCAACTCATACCTGACA- 298
DB 2918 TCCCATCTATTTGGCCAGGATTAAGCCCAAGACTTGAGCCAACTCATACCTGACA- 2977
QY 299 -CTTGCTCTTGGTAGGTGATGATTTTCTTTGGCCGCCCATTCAGAAAACCTTGTGCCA 357
DB 2978 TCTGTCTCTTGTAGGTGATGATTTTCTTTGGCCGCCCATTCAGAAAACCTTGTGCCA 3037
QY 358 TCAAGCACCAAGCGCTTTCATTTCTGCTACCTGTAGGCTACATGTTTCCAAAC 417
DB 3038 TCAAGCACCAAGCGCTTTCATTTCTGCTACCTGTAGGCTACATGTTTCCAAAC 3057
QY 418 AAGGCTCAACTGTGCTACAGCAGT---TACTTAGGGCTAAATTTATCCAAAGCAC 473
DB 3098 AAGGCTCAACTGTGCTACAGCAGT---TACTTAGGGCTAAATTTATCCAAAGCAC 3157
QY 474 CAGGCGCTCAGTGAAGAACATCCAG 501
DB 3158 CAGGCGCTCAGTGAAGAACATCCAG 3185

RESULT 9
AX027480 AX027480 7582 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 30 from Patent FR2788784.
DEFINITION
ACCESSION AX027480
VERSION AX027480.1 GI:10188444
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 7582)
AUTHORS Mallet, F., Voisset, C. and Paranhos, B.G.
JOURNAL Patent: FR 2788784-A 30 28-JUL-2000;
BIO MERIEUX (FR)
FEATURES
source location/Qualifiers
1..7582
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others
ORIGIN
Query Match 88.2%; Score 442; DB 6; Length 7582;
Best Local Similarity 91.3%; Pred. No. 1e-120;
Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;
QY 1 CAAGATCTCAGGATTAATCAAT-GAGGCTGTTGTCCTCTATAGCCAGCTAGACCC 59
DB 2678 CAAGATCTCAGGATTAATCAATGAGGCGCTGTCTTTATACCAAGCTAGACCC 2737
QY 60 TTATCTCTGCTGCTCCCAATACAGAGAGACAGAGTGGTTTACAGTCTGACCTTCA 119
DB 2738 TTATCTCTGTTGTTCCCAATACAGAGAGACAGAGTGGTTTACAGTCTGACCTTMA 2797
QY 120 GGATCCCTTCTTCTGATCCCTGTAATCTGATCTGATCTGATCTGATCTGATCTG 179
DB 2798 GGATCCCTTCTTCTGATCCCTGTAATCTGATCTGATCTGATCTGATCTGATCTG 2857
QY 180 TACTTCAACCCCAATCACTCACTGACCTA-TTACCCCAAGGTTTCAAGGATAG 239
DB 2858 TACTTCAACCCCAATCACTCACTGACCTA-TTACCCCAAGGTTTCAAGGATAG 2917
QY 240 TCCCATCTATTGGCCAGGATTAAGCCCAAGCTTGAACCAATCTCTATACCTGAGCA- 298
DB 2918 TCCCATCTATTGGCCAGGATTAAGCCCAAGCTTGAACCAATCTCTATACCTGAGCA- 2977
QY 299 -CTTGTCTTGGTGGTGGATGATTTACTTTGGCCGCCATTCAGAAACCTTGTGCA 357
DB 2978 TCTTGTCTTGGTGGTGGATGATTTACTTTGGCCGCCATTCAGAAACCTTGTGCA 3037
QY 358 TCAAGCCACCCCAAGGCTTCAATTTCTCGCTAC CTGAGCTACATGTTTCAAAAC 417
DB 3038 TCAAGCCACCCCAAGGCTTCAATTTCTCGCTAC CTGAGCTACATGTTTCAAAAC 3097
QY 418 AAAGCTCAACTCTGCTCAGCAGAGT---TACTTAGGGCTAAATATCCAAAGCAC 473
DB 3098 AARACTARCTCTGCTCAGCAGAGTAAATAGTAAATATATCCAAAGCAC 3157
QY 474 CAGGGCCCTCAGTGAAGAACATCCAG 501
DB 3158 CAGGGCCCTCAGTGAAGAACATCCAG 3185

RESULT 10
AF045450/c AF045450 40205 bp DNA linear PRI 20-MAR-1998
LOCUS Homo sapiens chromosome 21q22.3 comid Q11M5, complete sequence.
DEFINITION
ACCESSION AF045450
VERSION AF045450.1 GI:2895783
KEYWORDS
HTG.
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 40205)
AUTHORS Taudien, S. and Rosenthal, A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40205)
AUTHORS Taudien, S., Nordstiek, G., Dagand, E., Hildmann, T., Drescher, B.,
Weber, J., Rosenthal, A. and Yaspo, M.L.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1998) Genome Analysis, Institut for Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source location/Qualifiers
1..40205
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="cosmid Q11M5"
repeat_region complement (293..586)
/evidence=not_experimental
/rpt_family="AluSc"
repeat_region complement (643..964)
/evidence=not_experimental
/rpt_family="MERVL"
exon complement (1053..1176)
/note="Genscan, score = 5.19%, comment = Internal_exon 124
bp frame: 2 phase: 1"
/evidence=not_experimental
repeat_region complement (1251..1859)
/evidence=not_experimental
/rpt_family="HERVL"
exon complement (2296..3220)
/evidence=not_experimental
/rpt_family="HERV16"
2818..2889
/note="Xpound exon prediction, score = 82% (0%)"
/evidence=not_experimental
complement (3221..3658)
/evidence=not_experimental
/rpt_family="MLT1C"
3359..3395
/note="Xpound exon prediction, score = 84% (0%)"
/evidence=not_experimental
complement (3874..4176)
/evidence=not_experimental
/rpt_family="AluDb"
complement (4330..4433)
/evidence=not_experimental
/rpt_family="L2"
4385..4638
/note="GRAIL, score = 95.000%, comment = excellent shadow"
/evidence=not_experimental
complement (5113..7899)
/evidence=not_experimental
/rpt_family="L1p14"
complement (6119..6598)
/note="GRAIL, score = 63.000%, comment = good"
/evidence=not_experimental
complement (7927..8045)
/evidence=not_experimental
/rpt_family="FLAM_A"
8048..8178
/evidence=not_experimental
/rpt_family="L2"
8771..8876
/note="MZRF, score = 67.4%"
/evidence=not_experimental
complement (8809..9187)
/evidence=not_experimental
/rpt_family="MERIA"
8837..8876
/note="GRAIL, score = 55.000%, comment = good shadow"
/evidence=not_experimental
complement (9113..9213)
exon

```
/note="GRAIL, score = 55.000%, comment = good"
/exon /evidence=not experimental
/repeat_region complement(9150. .9258)
/exon /evidence=not experimental
/repeat_region /rpt_family="LIMC3"
/exon 9259. .9647
/exon /evidence=not experimental
/repeat_region /rpt_family="LIMC3"
/exon complement(10138. .10427)
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT2E"
/exon 10662. .10875
/exon /evidence=not experimental
/repeat_region /rpt_family="MER58A"
/exon complement(11072. .1156)
/exon /note="GRAIL, score = 86.000%, comment = excellent"
/exon /evidence=not experimental
/repeat_region complement(11072. .11129)
/exon /note="Xpound exon prediction, score = 89% (0%)"
/repeat_region /evidence=not experimental
/exon 11478. .11597
/exon /evidence=not experimental
/repeat_region /rpt_family="L2"
/exon 11570. .11704
/exon /note="M2EF, score = 66.1%"
/exon /evidence=not experimental
/repeat_region /note="GRAIL, score = 99.000%, comment = excellent shadow"
/exon 11570. .11704
/exon /evidence=not experimental
/repeat_region complement(11687. .11735)
/exon /evidence=not experimental
/repeat_region /rpt_family="MIR"
/exon complement(11774. .12320)
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1E"
/exon complement(11828. .11949)
/exon /note="GRAIL, score = 50.000%, comment = good"
/repeat_region /evidence=not experimental
/exon 12623. .12843
/exon /evidence=not experimental
/repeat_region /rpt_family="AlusX"
/exon 13102. .13457
/exon /evidence=not experimental
/repeat_region /rpt_family="THB1B"
/exon complement(13481. .13519)
/exon /note="GRAIL, score = 65.000%, comment = good"
/repeat_region /evidence=not experimental
/exon 15065. .15188
/exon /evidence=not experimental
/repeat_region /rpt_family="HERVL"
/exon 15130. .15378
/exon /note="GC score = 10.60 (249bp); Region: GC content"
/misc_feature /evidence=not experimental
/repeat_region 15369. .16268
/exon /evidence=not experimental
/repeat_region /rpt_family="MERVL"
/exon complement(16623. .16714)
/exon /note="GRAIL, score = 72.000%, comment = good"
/exon /evidence=not experimental
/exon 16661. .16730
/exon /note="M2EF, score = 93.5%"
/exon /evidence=not experimental
/repeat_region 16880. .17318
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1C"
/exon 17326. .17436
/exon /evidence=not experimental
/repeat_region /rpt_family="LTR16C"
/exon complement(17540. .17675)
/exon /evidence=not experimental
/repeat_region /rpt_family="MIR"
/exon complement(17805. .17864)
/exon /evidence=not experimental
```

```
/rpt_family="MLT1C"
/exon 18076. .18098
/exon /note="GRAIL, score = 45.000%, comment = marginal shadow"
/exon /evidence=not experimental
/exon complement(18219. .18349)
/exon /note="GenScan, score = 2.99%, comment = Internal_exon 131
bp frame: 1 phase: 2"
/repeat_region /evidence=not experimental
/exon complement(18717. .19211)
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1C"
/exon complement(19252. .19399)
/exon /note="GRAIL, score = 71.000%, comment = good"
/exon /evidence=not experimental
/repeat_region 19269. .19350
/exon /note="GRAIL, score = 81.000%, comment = excellent shadow"
/exon /evidence=not experimental
/exon complement(19379. .19399)
/exon /note="Xpound exon prediction, score = 60% (0%)"
/exon /evidence=not experimental
/exon complement(19501. .19881)
/exon /note="M2EF, score = 50.6%"
/repeat_region /evidence=not experimental
/exon 19945. .20020
/exon /note="homology = 100.00%, score = 38, counts = 2"
/repeat_region /evidence=not experimental
/exon /rpt_type=tandem
/repeat_region /rpt_unit=tctctgcacattctctgtgagagtgctggttatcgc
/repeat_region 20223. .20643
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1E"
/exon 20645. .20910
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1F"
/exon 20961. .21007
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1F"
/exon 21483. .21781
/exon /evidence=not experimental
/repeat_region /rpt_family="AluDb"
/exon complement(21504. .21756)
/exon /note="M2EF, score = 79.4%"
/exon /evidence=not experimental
/repeat_region complement(22325. .22806)
/exon /evidence=not experimental
/repeat_region /rpt_family="MLT1C"

Query Match 86.8%; Score 435; DB 9; Length 40205;
Best Local Similarity 93.9%; Pred. No. 1.1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
```

```
QY 1 CAAGATCTCAGATTACATCAGAGCGTGTTCCTTATAGCCAGCTGACCTAGCCCT 60
DB 30626 CAGATCTCAGATTACATCAGAGCGTGTTCCTTATAGCCAGCTGACCTAGCCCT 30567
QY 61 TATACCTGCTTCCCAATAATACAGAGAGAGAGTGTTCAGCTCGACCTCAG 120
DB 30566 TATACCTGCTTCCCAATAATACAGAGAGAGAGTGTTCAGCTCGACCTCAG 30507
QY 121 GATGCTTCTTGCATCCCTGTACATCTGACTGCAATCTTGTTCCTTGAAGAT 180
DB 30506 GATGCTTCTTGCATCCCTGTACATCTGACTGCAATCTTGTTCCTTGAAGAT 30447
QY 181 ACTTCAAAACCAACATCTCACTCACTGACATATTATCCCAAGGTTCAAGGATAGT 240
DB 30446 CTTTCAAAACCAACATCTCACTCACTGACATATTATCCCAAGGTTCAAGGATAGT 30387
QY 241 CCCCATATTATGCGCAGGATTAGCCCAAGACTTGAGCCATCTCATACCTGACACT 300
DB 30386 CCCCATATTATGCGCAGGATTAGCCCAAGACTTGAGCCATCTCATACCTGACACT 30327
QY 301 -TGTCTTGAGGTGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGTGCAT 358
```

Db	30326	CCTGTCCTTCAGTGCATGATGATTTACTTTAG-TCGCCCTTCAGAAACCTTGCCAT	30367	repeat_region	2629..2831	/evidence=not_experimental
Qy	359	CAAGCACCACCAAGGCTCTTCAATTTCTCGCTA-CTGAGCTACATGTTTCCAAACCA	418	repeat_region	3012..3122	/rpt_family="AluSx"
Db	30266	CAAGCCACCAAGGCTCTTAAATTTCTCTCACTA-CTGTGGCTACAGATTTCCAAACCA	30207	repeat_region	3123..3478	/evidence=not_experimental
Qy	419	AAGGCTCAACTCTGTCTCAGCAGAGTTA---CT-AGGCTTAAATTTCCAAAGGACC	474	repeat_region	3479..3618	/rpt_family="MLT1J"
Db	30206	AAGGCTAGCTCTGTCTCAGCAGGTTAATGCT-AGGCTTAAATTTCCAAAGTCAC	30147	repeat_region	3619..3751	/evidence=not_experimental
Qy	475	AGGGCCCTCAGTGAAGAACATCCAG	501	exon	3752..3840	/note="GRAIL, score = 65.000%, comment = good shadow"
Db	30146	AGGGCCCTCAGTGAAGAACGTATCCAG	30120	exon	3841..4071	/note="GRAIL, score = 42.000%, comment = marginal"
RESULT	11			exon	4072..4249	/evidence=not_experimental
LOCUS	AF121782/c	142742 bp	DNA	linear	PRI 02-FEB-1999	
DEFINITION	Homo sapiens chromosome 21q22.3	FAC 206A10	complete sequence.	exon	4250..4385	/note="Xpound exon prediction, score = 70% (0%)"
ACCESSION	AF121782			repeat_region	4386..4610	/evidence=not_experimental
VERSION	AF121782.1	GI:4210991		exon	4611..4751	/note="WZEF, score = 95.1%"
KEYWORDS	HTG.			repeat_region	4752..4819	/evidence=not_experimental
SOURCE	Homo sapiens.			exon	4820..4871	/note="WZEF, score = 95.1%"
ORGANISM	Homo sapiens			exon	4872..4919	/note="WZEF, score = 95.1%"
REFERENCE	1 (bases 1 to 142742)			exon	4920..5071	/note="Genscan, score = 1.83%, comment = Initial_exon 6 bp frame: 1 phase: 0"
AUTHORS	Taudien,S., Dagnan,E., Hildmann,T., Nordstiek,G., Drescher,B., Schatkevo,R., Weber,U., Schilling,M., Menzel,U., Yaspo,M.L. and Rosenthal,A.			repeat_region	5072..5347	/evidence=not_experimental
FEATURES	Source	location/Qualifiers		repeat_region	5348..5457	/evidence=not_experimental
		1..142742	/organism="Homo sapiens"	repeat_region	5458..5619	/evidence=not_experimental
		/db_xref="taxon:9606"		repeat_region	5620..5751	/evidence=not_experimental
		/chromosome="21"		repeat_region	5752..5819	/evidence=not_experimental
		/map="21q22.3"		repeat_region	5820..5919	/evidence=not_experimental
		/clone="PAC 206A10"		exon	5920..6071	/note="GRAIL, score = 45.000%, comment = marginal"
		3..118		exon	6072..6249	/note="GRAIL, score = 45.000%, comment = marginal"
		/evidence=not_experimental		repeat_region	6250..6419	/evidence=not_experimental
		/rpt_family="MIR"		repeat_region	6420..6571	/evidence=not_experimental
		complement(140..415)		repeat_region	6572..6751	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	6752..6919	/evidence=not_experimental
		/rpt_family="MLT2B"		repeat_region	6920..7071	/evidence=not_experimental
		complement(668..881)		exon	7072..7249	/evidence=not_experimental
		/evidence=not_experimental		exon	7250..7419	/evidence=not_experimental
		/rpt_family="MER58A"		exon	7420..7571	/note="GRAIL, score = 71.000%, comment = good"
		complement(1078..1162)		exon	7572..7751	/evidence=not_experimental
		/note="GRAIL, score = 86.000%, comment = excellent shadow"		exon	7752..7919	/evidence=not_experimental
		/evidence=not_experimental		exon	7920..8071	/note="GRAIL, score = 81.000%, comment = excellent shadow"
		1114..1403		repeat_region	8072..8249	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	8250..8419	/note="homology = 100.00%, score = 38, counts = 2"
		/rpt_family="MLT11"		repeat_region	8420..8571	/evidence=not_experimental
		1481..1603		repeat_region	8572..8751	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	8752..8919	/rpt_unit="tctctgacacctctcgttgagaggtcggttaccgc
		/rpt_family="L2"		repeat_region	8920..9071	/rpt_type=tandem
		1576..1710		repeat_region	9072..9249	/evidence=not_experimental
		/note="WZEF, score = 66.1%"		repeat_region	9250..9419	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	9420..9571	/evidence=not_experimental
		1576..1710		repeat_region	9572..9751	/evidence=not_experimental
		/note="GRAIL, score = 99.000%, comment = excellent"		repeat_region	9752..9919	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	9920..10071	/evidence=not_experimental
		complement(1633..1741)		repeat_region	10072..10249	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	10250..10419	/evidence=not_experimental
		/rpt_family="MIR"		repeat_region	10420..10571	/evidence=not_experimental
		complement(1780..2326)		repeat_region	10572..10751	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	10752..10919	/evidence=not_experimental
		/rpt_family="MLT1B"		repeat_region	10920..11071	/evidence=not_experimental
		complement(1834..1955)		repeat_region	11072..11249	/evidence=not_experimental
		/note="GRAIL, score = 50.000%, comment = good shadow"		repeat_region	11250..11419	/evidence=not_experimental
		/evidence=not_experimental		repeat_region	11420..11571	/evidence=not_experimental

/note="MZF, score = 79.4%"
/evidence=not experimental
repeat_region complement(12346..12827)
/evidence=not experimental
/rpt_family="MLT1C"
12802..13180
/evidence=not experimental
/rpt_family="Alusg"
complement(13057..13161)
/note="MZF, score = 76.1%"
/evidence=not experimental
repeat_region complement(13226..13364)
/evidence=not experimental
/rpt_family="L1MC2"
complement(13613..13908)
/evidence=not experimental
/rpt_family="Alusx"
complement(14317..14614)
/evidence=not experimental
/rpt_family="L1R37A"
complement(14962..15036)
/note="MZF, score = 93%"
/evidence=not experimental
repeat_region 15295..15299
/evidence=not experimental
/rpt_family="L1PA6"
complement(15469..15497)
/evidence=not experimental
/note="Xpound exon prediction, score = 62% (0%)"
exon 16045..16768
/evidence=not experimental
repeat_region /evidence=not experimental
/rpt_family="MER21A"
complement(16555..16555)
/note="GRAIL, score = 54.000%, comment = good"
/evidence=not experimental
exon 16823..16850
/evidence=not experimental
repeat_region /evidence=not experimental
/rpt_family="MER34"
complement(16851..17208)
/evidence=not experimental
/rpt_family="THER1A"
17209..17632
/evidence=not experimental
repeat_region /evidence=not experimental
/rpt_family="MER34"
complement(17633..17669)
/evidence=not experimental
/rpt_family="Alusg1"
complement(18004..18334)
/evidence=not experimental
repeat_region /evidence=not experimental
/rpt_family="L1R17"
complement(18335..18632)
/evidence=not experimental
/rpt_family="HERV17"
18522..18420
/note="GRAIL, score = 51.000%, comment = good shadow"
/evidence=not experimental
exon complement(18397..18584)
/note="GRAIL, score = 66.000%, comment = good"
/evidence=not experimental
repeat_region complement(18643..24647)
/evidence=not experimental
/rpt_family="HERV17"
complement(19228..19361)
/note="MZF, score = 92.8%"
/evidence=not experimental
exon 20900..21135
/note="MZF, score = 51.4%"
/evidence=not experimental
exon 21150..21275
/note="Xpound exon prediction, score = 74% (0%)"
/evidence=not experimental
Query Match 86.8%; Score 435; DB 9; Length 142742;

Best Local Similarity 93.9%; Pred. No. 1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAAGTCGAGGATTAATCAATGAGGCGTGTTCCTATATGACGAGCTGATACCTGACCT 60
Db 20632 CAAAGTCGAGGATTAATCAATGAGGCGTGTTCCTATATGACGAGCTGATACCTGACCT 20573
QY 61 TATACCTGCTTCCCAATATCCAGAGAGAGAGGTTTACAGTCTGACCTGACCT 120
Db 20572 TATACCTGCTTCCCAATATCCAGAGAGAGAGGTTTACAGTCTGACCTGACCT 20513
QY 121 GATGCTTCTTTCGATCCCTGATACATCTGACTCTGCAATCTTGTTCCTTGAAGAT 180
Db 20512 GATGCTTCTTTCGATCCCTGATACATCTGACTCTGCAATCTTGTTCCTTGAAGAT 20453
QY 181 ACTTCAACCCCAATCTCAATCACTGACATATTTTACCCCAAGGCTCAGGATAGT 240
Db 20452 CTTTCAACCCCAATCTCAATCACTGACATATTTTACCCCAAGGCTCAGGATAGT 20393
QY 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCCCTATACCTGACACT 300
Db 20392 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCCCTATACCTGACACT 20333
QY 301 -TGTCTTCGGTAGGTGATGATTTACTTTTGCCGCCCATTCAGAACTTGTGCAT 358
Db 20332 CTTGTCTTCGGTAGGTGATGATTTACTTTTGCCGCCCATTCAGAACTTGTGCAT 20273
QY 359 CAAAGCCCAAGGCTCTTCAATTTCTGCTACCTGTGTGCTACAGTGTTCACCAACCA 418
Db 20272 CAAAGCCCAAGGCTCTTCAATTTCTGCTACCTGTGTGCTACAGTGTTCACCAACCA 20213
QY 419 AAGGCTCACTGCTGCTACAGCAGGTTA---CTTAGGCTTAAATATTCACCAAGCACC 474
Db 20212 AAGGCTCACTGCTGCTACAGCAGGTTAATGCTTAGGGCTTAAATATTCACCAAGTACC 20153
QY 475 AAGGCTCTCAGTGAAGAACATCCAG 501
Db 20152 AAGGCTCTCAGTGAAGAACATCCAG 20126
RESULT 12
HS21C080/c 340000 bp DNA linear PRI 24-MAY-2000
LOCUS HS21C080
DEFINITION Homo sapiens chromosome 21 segment HS21C080.
ACCESSION AL163280 AP001735 BA000005
VERSION AL163280.2 GI:7717369
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polly,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schliebel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Norisaki,G.,
Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Rieseemann,L., Dagand,F., Wehrmeyer,S., Borzym,K., and
Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
Yaspo,M.L.
Direct Submision
Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
Group * Institute of Molecular Biotechnology, Genome Analysis *
Keio University School of Medicine, Dept. of Molecular Biology *
GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
Genetics (addresses see below)
The Chromosome 21 Mapping and Sequencing Consortium consists of
COMMENT


```
/rpt_type=DISPERSED
complement(11315..11556)
/note="AluJo"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
11702..11753
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
12649..12678
/gene="SH3BGR"
/number=5
complement(13115..13425)
/note="AluSx"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(13598..13892)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
14013..14131
/note="(CA)n"
/rpt_family="Simple_repeat"

Query Match      86.8%; Score 435; DB 9; Length 34000;
Best Local Similarity 93.9%; Pred. No. 1e-118;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGATTATCATATGAGGCTGTTCTCTATATGACGCTGTAACCTGACCCCT 60
    |||
Db 244792 CAAGATCTCAGATTATCATATGAGGCTGTTCTCTATATGACGCTGTAACCTGACCCCT 244733

QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
    |||
Db 244732 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244673

QY 121 GATGCTCTCTTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCAT 180
    |||
Db 244672 GATGCTCTCTTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCAT 244613

QY 181 ACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCCAAGGCTCAGGAGATAGT 240
    |||
Db 244612 CATTCAAAACCAATCTCACTCACTGAGCTATTTTACCCCAAGGCTCAGGAGATAGT 244553

QY 241 CCCCATCTATTTGGCCAGGAGATTAGCCCAAGCTGAGCAATCTGATCCTGAGACAT 300
    |||
Db 244552 CCCCATCTATTTGGCCAGGAGATTAGCCCAAGCTGAGCAATCTGATCCTGAGACAT 244493

QY 301 --TGTCCTTGAGTATGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGCCAT 358
    |||
Db 244492 CTTGTCCTTGAGTATGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGCCAT 244433

QY 359 CAAGCCACCAAGCGCTTCAATTTCTGCTACCTGTCGCTACATGTTTCCAAACCA 418
    |||
Db 244432 CAAGCCACCAAGCGCTTCAATTTCTGCTACCTGTCGCTACATGTTTCCAAACCA 244373

QY 419 AAGGCTCACTGCTGCTCAGCAGGTTA-----CTTAGGGCTAAATATTCAGAGGAC 474
    |||
Db 244372 AAGGCTCACTGCTGCTCAGCAGGTTAAGCTTAGGGCTAAATATTCAGAGGAC 244313

QY 475 AGGGCCCTCAGTGAAGAAACATCCAG 501
    |||
Db 244312 AGGGCCCTCAGTGAAGAAACATCCAG 244286

RESULT 13
LOCUS      AL162912      8339 bp      DNA      linear      PRI 22-NOV-2001
DEFINITION Human DNA sequence from clone RP6-198C4 on chromosome Xq13.3-21.2,
ACCESSION  AL162912 AL121824
VERSION     AL162912.1 GI:7406722
KEYWORDS    HTG.
```

```
SOURCE      human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 8339)
AUTHORS      Pearce, A.
TITLE        Direct Submission
JOURNAL      Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em; EMBL; Sw; SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrx
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. RP6-198C4 is from the
library RPL1-6 constructed by the group of Pieter de Jong. For
further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pPAC4
IMPORTANT: This sequence is not the entire insert of clone
RP6-198C4. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
FEATURES     location/Qualifiers
source       1..8339
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="X"
             /map="q13.3-21.2"
             /clone="RP6-198C4"
             /clone_11b="RPL1-6"
             1..5641
             /note="HERV17 repeat: matches 368..5930 of consensus"
             repeat_region 5635..5673
             /note="HERV17 repeat: matches 7419..7457 of consensus"
             repeat_region 5664..5955
             /note="HERV17 repeat: matches 8244..8523 of consensus"
             repeat_region 5956..6136
             /note="LTR17 repeat: matches 1..239 of consensus"
             repeat_region 6195..6280
             /note="LTR17 repeat: matches 239..324 of consensus"
BASE COUNT  2649 a 1846 c 1726 g 2118 t
ORIGIN
Query Match      85.5%; Score 428.6; DB 9; Length 8339;
Best Local Similarity 93.1%; Pred. No. 9.8e-117;
Matches 472; Conservative 0; Mismatches 29; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGATTATCATATGAGGCTGTTCTCTATATGACGCTGTAACCTGACCCCT 60
    |||
Db 3668 CAAGATCTCAGATTATCATATGAGGCTGTTCTCTATATGACGCTGTAACCTGACCCCT 3727

QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
    |||
Db 3728 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3787

QY 121 GATGCTCTCTTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCATCCTGTCAT 180
```


Mon Apr 21 09:59:02 2003

us-09-719-554-3_copy_5000_5500.rge

Page 15

Search completed: April 19, 2003, 13:51:57
Job time : 1888.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 10:05:38 ; Search time 180.12 Seconds

(without alignments)
6263.892 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501

Sequence: 1 caagatctcagatataca.....tcagtggagacacatccag 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	21	ABN97929
2	501	100.0	56093	24	ABL61744
3	474	94.6	2938	20	AA525656
4	474	94.6	2938	21	AA59206
5	474	94.6	7466	23	AA568626
6	453.8	90.6	3831	23	AA571727
7	442	88.2	7582	20	AA525655
8	442	88.2	7582	21	AA59215
9	435	86.8	3867	23	AA568621

10	435	86.8	3867	23	AA576464	DNA encoding novel
11	435	86.8	3867	23	AA580471	DNA encoding novel
12	435	86.8	4349	23	AA576475	DNA encoding novel
13	435	86.8	4535	23	AA576205	DNA encoding novel
14	435	86.8	5154	23	AA567609	DNA encoding novel
15	428.4	85.5	3372	20	AA525663	Human endogenous r
16	428.4	85.5	3372	21	AA59213	Partial pol gene a
17	422.2	84.3	900	23	AA567599	DNA encoding novel
18	422.2	84.3	900	23	AA572228	DNA encoding novel
19	422.2	84.3	900	23	AA573911	DNA encoding novel
20	422.2	84.3	900	23	AA576470	DNA encoding novel
21	422.2	84.3	1483	23	AA567606	DNA encoding novel
22	422.2	84.3	1687	23	AA572234	DNA encoding novel
23	422.2	84.3	1729	23	AA577018	DNA encoding novel
24	422.2	84.3	1729	23	AA584207	DNA encoding novel
25	422.2	84.3	2304	19	AA543199	Multiple sclerosis
26	420.6	84.0	924	23	AA572224	DNA encoding novel
27	420.6	84.0	1660	23	AA573608	DNA encoding novel
28	420.6	84.0	1660	23	AA576459	DNA encoding novel
29	420.6	84.0	1660	23	AA584201	DNA encoding novel
30	420.6	84.0	1810	23	AA585801	DNA encoding novel
31	420.6	84.0	1810	23	AA586023	DNA encoding novel
32	420.6	84.0	1846	23	AA573605	DNA encoding novel
33	420.6	84.0	1846	23	AA584192	DNA encoding novel
34	420.6	84.0	2220	23	AA567604	DNA encoding novel
35	420.6	84.0	2220	23	AA573919	DNA encoding novel
36	420.6	84.0	2220	23	AA576473	DNA encoding novel
37	419.6	83.8	1158	16	AA5702815	Multiple sclerosis
38	419.6	83.8	1158	17	AA573773	Partial sequence o
39	419.6	83.8	1158	18	AA576527	MSRV-1B pol* seque
40	419.6	83.8	1158	19	AA543115	Multiple sclerosis
41	419.6	83.8	2365	19	AA543200	Multiple sclerosis
42	419.6	83.8	2365	19	AA576478	MSRV-1 pol gene se
43	419.6	83.8	2391	19	AA543166	Multiple sclerosis
44	419	83.6	1117	23	AA576460	DNA encoding novel
45	419	83.6	1389	23	AA567594	DNA encoding novel

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
XX
AC ABN97929;
XX
DT 01-AUG-2002 (first entry)
XX
DE Human retroviral sequence HERV-7g.
XX
KW Autoimmune disease; HERV-7g; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX
OS Human retrovirus.
XX
PN WO9967395-A1.
XX
PD 29-DEC-1999.
XX
PF 23-JUN-1999; 99WO-FR01513.
XX
PR 23-JUN-1998; 98FR-0007920.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Alliel PM, Perin J, Rieger F;
XX
DR WPI; 2000-160587/14.
XX
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225pp; French.
XX
PS
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 1,6e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTTATAGCCAGCTGTACTAGCCCT 60
DB 5000 CAAGATCTCAGGATTATCATGAGGCTGTGTTCTTATAGCCAGCTGTACTAGCCCT 5059
QY 61 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 5060 TATATCTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY 121 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 5120 GATGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5179
QY 181 ACTTCAAAACCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 240
DB 5180 ACTTCAAAACCAATCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 5239
QY 241 CCCCACTATTGGGCGAGGATTAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 5240 CCCCACTATTGGGCGAGGATTAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5299
QY 301 TGTCTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 5300 TGTCTTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5359
QY 361 AGCCACCAAGGCTCTTCAATTTCTGCTAAGCTTGGCTACATGTTTCCAAACCAA 420
DB 5360 AGCCACCAAGGCTCTTCAATTTCTGCTAAGCTTGGCTACATGTTTCCAAACCAA 5419
QY 421 GGCTCAACTGTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 480
DB 5420 GGCTCAACTGTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 5479
QY 481 CTCAGTGAAGAACATCCAG 501
DB 5480 CTCAGTGAAGAACATCCAG 5500

KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-23133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237315P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
(AVAL-) AVALON PHARM.
PA
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Sopet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
XX chemical agent to be tested for anti-neoplastic activity, and
XX determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 81; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (I)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (I) has cytosstatic

AA59206,
07-NOV-2000 (first entry)
Gag and partial pol gene fragment of HERV-W from human genome.
Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
Homo sapiens.
MO200043521-A2.
27-JUL-2000.
21-JAN-2000; 2000WO-FR00144.
21-JAN-1999; 99FR-0000888.
(INMR) BIO MERIEUX.
Paranhos-Baccala G, Mallet F, Voisset C;
WPI; 2000-499229/44.
New nucleic acid from human endogenous retrovirus, useful e.g. for
diagnosis of autoimmune disease and complications of pregnancy,
contains at least part of the gag gene -
Disclosure; Page 43; 53pp; French.
The present sequence represents an endogenous retroviral nucleic acid
fragment, which is associated with an autoimmune disease, and is
integrated into the human genome. The fragment is originally derived
from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
HERV-W retrovirus is associated with autoimmune disease, failure of
pregnancy or disorders of pregnancy. The nucleic acid fragment, or
proteins derived from it, are useful for diagnosis of autoimmune
disease (specifically multiple sclerosis) and for monitoring pregnancy.
The nucleic acid fragments may also be used for in situ labelling of
isolated chromosomes, while the transcription product can be used to
study or monitor T cell proliferation in vitro.
Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;
Query Match 94.6%; Score 474; DB 21; Length 2938;
Best Local Similarity 97.8%; Pred. No. 5=143;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 CAAATCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGTAACCTAGCC 59
DB 2079 CAAATCTCAGGATTATCAATGAGGCGCTGTCTTATATCCAGCTGTAACCTAGCC 2138
QY 60 TTATATCTGCTTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
DB 2139 TTATATCTGCTTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
QY 120 GGATGCGCTTCTTCCGACCCGCTGATCCGACTCTCAATTTCTGTTGCTTGAAGA 179
DB 2199 GGAATGCGCTTCTTCCGACCCGCTGATCCGACTCTCAATTTCTGTTGCTTGAAGA 2258
QY 180 TACTTCAAACCAATCTCAACTCACTGACCTGACTATTTTACCCCAAGGTTGAGGATAG 239
DB 2259 TACTTCAAACCAATCTCAACTCACTGACCTGACTTTTACCCCAAGGTTGAGGATAG 2318
QY 240 TCCCATCTATTTGGCCAGGCTTATGCCCAAGACTTATAGCCAACTCTATATCTGAGAC 299
DB 2319 CCCCCTATATTTGGCCAGGCTTATGCCCAAGACTTATAGCCAACTCTATATCTGAGAC 2378
QY 300 TTGTCCTCGGAGGTGATGATTACTTTTGGCCGACCATTCAGAAACCTTGTGCCATC 359
DB 2379 TTGTCCTCGGAGGTGATGATTACTTTTGGCCGACCATTCAGAAACCTTGTGCCATC 2438

QY 360 AAGCCACCAAGGCTCTTAATTTCTCGCTACCTGTGCTACATGTTTCCAAACCA 419
DB 2439 AAGCCACCAAGGCTCTTAATTTCTCGCTACCTGTGCTACATGTTTCCAAACCA 2498
QY 420 AGGCTCAACTCTGTCTACACAGGTTACTTATAGGCTTAAATTTTCCAAAGGACACAGGC 479
DB 2499 AGGCTCAACTCTGTCTACACAGGTTACTTATAGGCTTAAATTTTCCAAAGGACACAGGC 2558
QY 480 CCTCAGTAGGAACACATCCAG 501
DB 2559 CCTCAGTAGGAACACATCCAG 2580
RESULT 5
AAS68626
ID AAS68626 standard; cDNA; 7466 BP.
AC AAS68626;
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #4430.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX MO200175067-A2.
XX PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR P-PSDB; ABG04439.
XX DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 4430; 103pp; English.
XX PS The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations in
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;

Query Match 94.6%; Score 474; DB 23; Length 7466;

Best Local Similarity 97.8%; Pred. No. 7, 6e-143;

Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```
QY 1 CAAGATCTCAGATTCATCAT-GAGGCTGTTGTTCTCTATAGCCAGCTGACTGACCC 59
DB 6607 CAAGATCTCAGATTCATCATGAGGCGCTTGTCCTTTATACCCAGCTGACTGACCC 6666
QY 60 TTATACCTGCTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
DB 6667 TTATACCTGCTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6726
QY 120 GAGTGCCTTCTTGCATCCCTGATCATCTGATCTCAATCTTGTGCTTGAAGA 179
DB 6727 GAGTGCCTTCTTGCATCCCTGATCATCTGATCTCAATCTTGTGCTTGAAGA 6786
QY 180 TACTTCAAAACCAACATCTCACTGACCTGACTATTTTACCACCAAGGTTCAAGGATAG 239
DB 6787 TACTTCAAAACCAACATCTCACTGACCTGACTATTTTACCACCAAGGTTCAAGGATAG 6846
QY 240 TCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCACTCTATACCTGACAC 299
DB 6847 CCCCACATATTTGGCCAGGATTAAGCCCAAGCTTGAAGCACTCTATACCTGACAC 6906
QY 300 TTGCTCTCGGTAGTGTGATGATTAATTTTGGCCGCTTGAAGAAACCTTGACATC 359
DB 6907 TTGCTCTCGGTAGTGTGATGATTAATTTTGGCCGCTTGAAGAAACCTTGACATC 6966
QY 360 AAGCCACCCAGGCGCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 419
DB 6967 AAGCCACCCAGGCGCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 7026
QY 420 AGGCTCAACTCTGCTCAACAGAGTTACTTGAAGCTTAAATTTCCAAAGCACAGGCG 479
DB 7027 AGGCTCAACTCTGCTCAACAGAGTTACTTGAAGCTTAAATTTCCAAAGCACAGGCG 7086
QY 480 CCTCAGTGAAGAACATCCAG 501
DB 7087 CCTCAGTGAAGAACATCCAG 7108
```

RESULT 6

AA571727
ID AA571727 standard; cDNA; 3831 BP.

XX AA571727;

AC AA571727;

DE 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #7531.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG07540.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID No 7531; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AA564197-AA594564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3831 BP; 1173 A; 953 C; 815 G; 890 T; 0 other;

Query Match 90.6%; Score 453.8; DB 23; Length 3831;

Best Local Similarity 95.2%; Pred. No. 2e-136; Mismatches 22; Indels 2; Gaps 1;

Matches 479; Conservative 0; Mismatches 22; Indels 2; Gaps 1;

```
QY 1 CAAGATCTCAGATTCATCATGAGGCGTGTGCTCTATAGCCAGCTGACTGACCTCCT 60
DB 412 CAAGATCTCAGATTCATCATGAGGCGTGTGCTCTATAGCCAGCTGACTGACCTCCT 471
QY 61 TATACCTGCTTTCCTCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 472 TATACCTGCTTTCCTCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
QY 121 GATGCTTCTGCTGATCCCTGATCATCTGATCTCAATCTTGTGCTTGAAGAT 180
DB 532 GATGCTTCTGCTGATCCCTGATCATCTGATCTCAATCTTGTGCTTGAAGAT 591
QY 181 ACTTCAAAACCAACATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGT 240
DB 592 CTTCAAAACCAACATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGT 651
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAAGCACTCTCATCCTGACAC 298
DB 652 CTTCAAAACCAACATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGT 711
QY 299 CTTGCTCTCGGTAGTGTGATGATTAATTTTGGCCGCTTGAAGCACTCTCATCCTGACAT 358
DB 712 CTTGCTCTCGGTAGTGTGATGATTAATTTTGGCCGCTTGAAGCACTCTCATCCTGACAT 771
QY 359 CAAGCCACCAAGCGCTTTCATATTTCTCGCTACCTGTGCTACATGTTTCCAAACCA 418
DB 772 CAAGCCACCAAGCGCTTTCATATTTCTCGCTACCTGTGCTACATGTTTCCAAACCA 831
QY 419 AAGGCTCAACTCTGCTCAACAGAGTTACTTGAAGGCTTAAATTTATCCAAAGGCAACAGG 478
DB 832 AAGGCTCAACTCTGCTCAACAGAGTTACTTGAAGGCTTAAATTTATCCAAAGGCAACAGG 891
QY 479 CCTCAGTGAAGAACATCCAG 501
DB 892 CCTCAGTGAAGAACATCCAG 914
```

ID	AXX25665	standard; cDNA to mRNA; 7582; BP.
XX	AXX25665;	
XX	21-MAY-1999	(first entry)
XX	Complete human endogenous retrovirus W genome.	
XX	Clone; human endogenous retrovirus; genome; autoimmune disease;	
XX	multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;	
XX	disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.	
XX	Human endogenous retrovirus.	
XX	MO9902696-A1.	
XX	21-JAN-1999.	
XX	06-JUL-1998;	98WO-FR01442.
XX	07-JUL-1997;	97FR-0008815.
XX	(IMMR) BIO MERIEUX.	
XX	Beseme F, Blond JL, Bouton O, Malle F, Mandrand B;	
XX	WPI; 1999-120897/10.	
XX	New nucleic acid sequences from human endogenous retrovirus-W -	
XX	expressed exclusively in placenta and useful in diagnosis and	
XX	therapy of autoimmune disease, and abnormal or failed pregnancy	
XX	Claim 1; Page 71-74; 106pp; French.	
XX	This sequence represents the complete sequence of the human endogenous	
XX	retrovirus (HERV) W genome. The nucleic acids, their fragments or	
XX	peptides encoded by them are markers of autoimmune disease (e.g. multiple	
XX	sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,	
XX	insulin-dependent diabetes and related pathologies) and of abnormal or	
XX	unsuccessful pregnancy and can be used as chromosomal markers for	
XX	susceptibility to these conditions, or proximity markers of genes	
XX	associated with this susceptibility.	
XX	Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;	
XX	Query Match 88.2%; Score 442; DB 20; Length 7582;	
XX	Best Local Similarity 91.3%; Pred. No.1.8e-132;	
XX	Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;	
QY	1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGTACGACCC	59
DB	2678 CAAGATCTCAGGATTATCAATGAGGCGCTGTCTTATATCCGCGTGTACTGACCC	2737
QY	60 TTATATCTGCTTCCCAATATCCAGAGAGAGAGGTGTTTACAGTCTGAGACCTTCA	119
DB	2738 TTATATCTGATGTTTCCCAATATCCAGAGAGAGAGGTGTTTACAGTCTGAGACCTTMA	2797
QY	120 GGATGCTTCTTCTGTGATCCCTGTACATCTGACTTCAATTTCTGTGTTGCTTTGAGA	179
DB	2798 GGATGCTTCTTCTGTGATCCCTGTACATCTGACTTCAATTTCTGTGTTGCTTTGAGA	2857
QY	180 TACTTCAACCCCAACTCTCAACTCACTCCGTGACTATTTTACCCCAAGGTTTACGGGATG	239
DB	2858 TACTTCAACCCCAACTCTCAACTCACTCCGTGACTATTTTACCCCAAGGTTTACGGGATG	2917
QY	240 TCCCATCATTTTGGCCAGAGCATTTAGCCCAAGACTTACCCCACTTACTGAGCA-298	
DB	2918 TCCCATCATTTTGGCCAGAGCATTTAGCCCAAGACTTACCCCACTTACTGAGCA-2977	
QY	299 -CTGTGCTTGGATGATGATTTACTTTTGGCTGCCCATTTACGAAACTTGTGCCA	357

Dn	2978	TCTTGTCCCTCCTGTAAGTGAGATTAACCTTTTRCGCYGCVRITTCAGAAACTTGTGCA	3037
Oy	358	TCAGACCACCCAAAGCCCTCTTCAATTTCCTCGCTACCTGTGCTPACATNGTTTTCCAAC	417
Dd	3038	TCAGGCCACCCCAAGCRCTCTTMAATTTCCTCGCYACCTGTGGCTAACAGGTTTCCAACS	3097
Oy	418	AAGGCTCAACTGTGTCACAGAGGT-----TACTTAGGGCTAATAATATCCAAAGGCAC	473
Dd	3098	ARAGCTCAACTGTGTCACAGAGGTTAATAATCTTAGGRTAARATATATCCAAAGGCAC	3157
Oy	474	CAGGGCCCTCAGTGAAGAACATCCAG	501
Dd	3158	CARGCCCTCAGTGAAGAAYRATCCAG	3185
<hr/>			
RESULT 8			
XX	AAA59215		
AC	AAA59215	standard; DNA; 7582 BP.	
XX	AAA59215;		
DE	07-NOV-2000	(first entry)	
XX			
XX	Human endogenous retrovirus W (HERV-W) sequence.		
KM	Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;		
KW	gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.		
XX			
OS	Human endogenous retrovirus.		
XX			
FT	Key	Location/Qualifiers	
FT	LTR	1..120	
FT		/tag= a	
FT		/note= "R of 5' LTR"	
FT	LTR	121..575	
FT		/tag= b	
FT		/note= "U5 of 5' LTR"	
FT	primer_bind	579..596	
FT		/tag= c	
FT	CDS	581..7194	
FT		/tag= d	
FT	CDS	7039..7194	
FT		/tag= e	
FT		/note= "ORF2 52 AA"	
FT	CDS	7112..7255	
FT		/tag= f	
FT		/note= "ORF3 48 AA"	
FT	misc_feature	7244..7254	
FT		/tag= g	
FT		/note= "polyurine tract"	
FT	LTR	7256..7582	
FT		/tag= h	
FT		/note= "U3-R of 3' LTR"	
FT	polyA_signal	7563..7569	
FT		/tag= i	
XX			
XX	MO200043521-A2.		
XX			
PD	27-JUL-2000.		
XX			
PE	21-JAN-2000; 2000MO-FR00144.		
XX			
PR	21-JAN-1999; 99FR-000888.		
XX			
PA	(IMMR) BIO MERIEUX.		
PI	Paranhos-Baccala G, Mallet F, Voisset C;		
XX			
DR	WPI; 2000-499229/44.		
XX			
PT	New nucleic acid from human endogenous retrovirus, useful e.g. for		
PT	diagnosis of autoimmune disease and complications of pregnancy,		

PT	contains at least part of the gag gene	-
XX	Disclosure: Page 49-52; 53pp; French.	
PS		
CC	The present sequence represents an endogenous retrovirus, which is	
CC	associated with an autoimmune disease, and is integrated into the human	
CC	genome. The retrovirus is human endogenous retrovirus W (HERV-W). The	
CC	HERV-W retrovirus is associated with autoimmune disease, failure of	
CC	pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or	
CC	proteins derived from it, are useful for diagnosis of autoimmune	
CC	disease (specifically multiple sclerosis) and for monitoring pregnancy.	
CC	The nucleic acid fragments may also be used for in situ labelling of	
CC	isolated chromosomes, while the transcription product can be used to	
CC	study or monitor T cell proliferation in vitro.	
XX		
XX	Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;	
SO		
	Query Match	88.2%; Score 442; DB 21; Length 7582;
	Best Local Similarity	91.3%; Pred. No. 1.8e-132;
	Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3	
QY	1 CAAGATCTCAGGATTATCAAT-GAGGCTGTTGTCCTCTATAGCCAGCTGTAAGCCC	59
Db	2678 CAAGATCTCAGGATTATCAATGAGGCGCTTGCTTTATACCGAGCTGTAAGCCC	2737
QY	60 TTAATCTGCTCTTCCCAATACAGAGAGCAAGAGTGTTTAAAGTCTGAGACTTCA	119
Db	2738 TTAATCTGCTGTTCCCAATATCCAGAGAGCAAGAGTGTTTAAAGTCTGAGACTTCA	2797
QY	120 GAGATCCCTTTCTTGCAATCCCTGTAATCCCTGTAATCTGTAATCTGTTGCTTTGAACA	179
Db	2798 GAGATCCCTTTCTTGCAATCCCTGTAATCCCTGTAATCTGTAATCTGTTGCTTTGAACA	2857
QY	180 TACTTCAACCAACCAATCTCACTCACTGACTATTTTAAACCAAGGTTGAGGATAG	239
Db	2858 TACTTCAACCAACCAATCTCACTCACTGACTATTTTAAACCAAGGTTGAGGATAG	2917
QY	240 TCCCATTTAATTGGCCAGGACATTAGCCCAAGACTTTAGCCAACTCTCATCTGAGAC-	298
Db	2918 TCCCATTTAATTGGCCAGGACATTAGCCCAAGACTTTAGCCAACTCTCATCTGAGAC-	2977
QY	299 -CTTGTCTTGGGTAGTGATGATTTACTTTTGGCCGCCCAATTCAGAAACCTTGAGCA	357
Db	2978 TCTTGTCTTGGGTAGTGATGATTTACTTTTGGCCGCCCAATTCAGAAACCTTGAGCA	3037
QY	358 TCAAGCAACCAAGGCTCTTCAATTTCTGCTGCTACCTGCTACATGTTTCCAAACC	417
Db	3038 TCAAGCAACCAAGGCTCTTCAATTTCTGCTGCTACCTGCTACATGTTTCCAAACC	3097
QY	418 AAAGGCTCAACTGCTGCTCAACAGCAGGT----TACTTAAAGGCTAAATTTCCAAAGGCAC	473
Db	3098 AARACTCAACTGCTGCTCAACAGCAGGTAAATTTTAAATTTTCCAAAGGCAC	3157
QY	474 CAGGCGCTCAGTGAGAACACATCCAG	501
Db	3158 CAGGCGCTCAGTGAGAACATCCAG	3185
RESULT 9		
AA58621	AA58621 standard; cDNA; 3867 BP.	
XX	AA58621;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	DNA encoding novel human diagnostic protein #4425.	
XX		
KM	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
XX	Homo sapiens.	

XX	WC0200175067-A2.
XX	11-OCT-2001.
XX	30-MAR-2001; 2001WO-US08631.
XX	31-MAR-2000; 2000US-0540217.
XX	23-AUG-2000; 2000US-0649167.
XX	(HYSE-) HYSEQ INC.
XX	Dzmanac RT, Liu C, Tang YT;
XX	WPI; 2001-639362/73.
XX	P-PSDB; ABG04434.
XX	New isolated polynucleotide and encoded polypeptides, useful in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits and to assess
XX	biodiversity -
XX	Claim 1; SEQ ID No 4425; 103pp; English.
XX	The invention relates to isolated polynucleotide (I) and
XX	polypeptide (II) sequences. (I) is useful as hybridization probes,
XX	polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome
XX	and gene mapping, and in recombinant production of (II). The
XX	polynucleotides are also used in diagnostics as expressed sequence tags
XX	for identifying expressed genes. (I) is useful in gene therapy techniques
XX	to restore normal activity of (II) or to treat disease states involving
XX	(II). (II) is useful for generating antibodies against it, detecting or
XX	quantitating a polypeptide in tissue, as molecular weight markers and as
XX	a food supplement. (II) and its binding partners are useful in medical
XX	imaging of sites expressing (II). (I) and (II) are useful for treating
XX	disorders involving aberrant protein expression or biological activity.
XX	The polypeptide and polynucleotide sequences have applications in
XX	diagnostics, forensics, gene mapping, identification of mutations
XX	responsible for genetic disorders or other traits to assess biodiversity
XX	and to produce other types of data and products dependent on DNA and
XX	amino acid sequences. AAS64197-AAS94564 represent novel human
XX	diagnostic coding sequences of the invention.
XX	Note: The sequence data for this patent did not appear in the printed
XX	specification, but was obtained in electronic format directly from WIPO
XX	at ftp.wipo.int/pub/published_pct_sequences.
XX	Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
XX	Query Match 86.8%; Score 435; DB 23; Length 3867;
XX	Best Local Similarity 93.9%; Pred. No. 2.5e-130;
XX	Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY	1 CAAGATCTCAGAGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGACTAGCCCT 60
DB	1693 CAAGATCTCAGAGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGACTAGCCCT 1752
QY	61 TATACTCTGCTTCCCAATACAGAGAGACAGAGTGGTTACAGTCTTGAGCTTGAG 120
DB	1753 TATACTCTGCTTCCCAATACAGAGAGACAGAGTGGTTACAGTCTTGAGCTTGAG 1812
QY	121 GATGCTTCTTCTGCATCCCTGTACATCCTGACTCTCAATCTTGTTGACCTTGAAGT 180
DB	1813 GATGCTTCTTCTGCATCCCTGTACATCCTGACTCTCAATCTTGTTGACCTTGAAGT 1872
QY	181 ACTTCAAAACCAATCTCAACTCACTGAGACTATTTTACCCCAAGGGTTACGGATAGT 240
DB	1873 ACTTCAAAACCAATCTCAACTCACTGAGACTATTTTACCCCAAGGGTTACGGATAGT 1932
QY	241 CCCCATCTATTGGCCAGGCGATTAGCCCAACACTTGAGCCCAATCCTCAADACTGGAGACT 300
DB	1993 CCCCATCTATTGGCCAGGCGATTAGCCCAACACTTGAGCCCAATCCTCAADACTGGAGACT 1992
QY	301 --TGTCTTCGGTAGTGATGATTTACTTTTGGCCGCCCATTCAGAAACCTGTGTCAT 358

Db 1993 CCTGTCTCTTCAGTCATGATGATTTACTTTAG:TTGCCCTTGAGAAACCTTGCCAT 2052
Qy 359 CAAGCCACCCAGGCGCTTCAATTTCTCGCTA:CTGTGGCTCATGTGTTCCAAACCA 418
Db 2053 CAAGCCACCCAGGCGCTTCAATTTCTCGCTA:CTGTGGCTCATGTGTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGTCTCAACAGAGTTA----CTTAGGCTTAAATTTCCAAAGGACC 474
Db 2113 AAGGCTCAACTCTGTCTCAACAGAGTTA:AGGCTTAAATTTATCCAAAGTACC 2172
Qy 475 AGGGCCCTCAGTAGGAACATCCAG 501
Db 2173 AGGGCCCTCAGTAGGAACGTATCCAG 2199

RESULT 10
AAS76464
ID AAS76464 standard; cDNA; 3867 BP.
XX
AC AAS76464;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12268.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KN food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI
DR WPI: 2001-639362/73.
DR P-PSDB; ABG12277.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 1; SEQ ID No 12268; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations in responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO

CC at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
XX
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No.2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
Qy 1 CAAGATCTCAGATATTAATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTAGCCCT 60
Db 1693 CAAGATCTCAGATATTAATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTAGCCCT 1752
Qy 61 TATATCTGCTTTTCCAAATACAGAGAGACAGAGTGTATACAGTCTTGACCTTACAG 120
Db 1753 TATATCTGCTTTTCCAAATACAGAGAGACAGAGTGTATACAGTCTTGACCTTACAG 1812
Qy 121 GATGCTCTTCTTGCAATCCCGTACATCTGACTCTCAATTTCTGTTGCTTGAAGAT 180
Db 1813 GATGCTCTTCTTGCAATCCCGTACATCTGACTCTCAATTTCTGTTGCTTGAAGAT 1872
Qy 181 ACTTCAAAACCCAAATCTCAATCACTGACCTGACATATTTTACCCCAAGGATTCAGGATAGT 240
Db 1873 CTTTCAAAACCCAAATCTCAATCACTGACCTGACATATTTTACCCCAAGGATTCAGGATAGT 1932
Qy 241 CCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCATCTCATACCTGACACT 300
Db 1933 CCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCATCTCATACCTGACACT 1992
Qy 301 --TGTCCTTGCTAGTGGAGATTTACTTTGGCCGCCCATTCAGAAACCTTGCCAT 358
Db 1993 CTTGCTCTTCAATGATGATGATTTACTTTTACCTGCTCCCTTCAAGAACCTTGCCAT 2052
Qy 359 CAAGCCACCCAGGCGCTTCAATTTCTCGCTA:CTGTGGCTCATGTGTTCCAAACCA 418
Db 2053 CAAGCCACCCAGGCGCTTCAATTTCTCGCTA:CTGTGGCTCATGTGTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGTCTCAACAGAGTTA----CTTAGGCTTAAATTTCCAAAGGACC 474
Db 2113 AAGGCTCAACTCTGTCTCAACAGAGTTA:AGGCTTAAATTTATCCAAAGTACC 2172
Qy 475 AGGGCCCTCAGTAGGAACATCCAG 501
Db 2173 AGGGCCCTCAGTAGGAACGTATCCAG 2199

RESULT 11
AAS80471
ID AAS80471 standard; cDNA; 3867 BP.
XX
AC AAS80471;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #16275.
XX
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KN food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI
XX

DR WPI, 2001-639362/73.
DR P-PSDB; ABG16284.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1; SEQ ID No 16275; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No. 2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCGAGTATCAATGAGAGCTGTCTCTCATGSCAGAGTATCCTAGCCCT 60
DB 1693 CAAGATCTCGAGTATCAATGAGAGCTGTCTCTCATGSCAGAGTATCCTAGCCCT 1752
QY 61 TATACCTGCTTTCCCAATACAGAGAGAGAGAGTGTATTAAGTCTTGACCTTCAG 120
DB 1753 TATACCTGCTTTCCCAATACAGAGAGAGAGAGTGTATTAAGTCTTGACCTTCAG 1812
QY 121 GATGCTTTCTTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
DB 1813 GATGCTTTCTTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 1872
QY 181 ACTTCAACCAACCAATGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 240
DB 1873 ACTTCAACCAACCAATGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 1932
QY 241 CCCCATGATTTGGCCAGGAGTATGAGCCAGTATGAGCCAGTATGAGCCAGTATGAG 300
DB 1933 CCCCATGATTTGGCCAGGAGTATGAGCCAGTATGAGCCAGTATGAGCCAGTATGAG 1992
QY 301 -TGTCTTGGTGAAGTATGATTTATCTTTGGCCGCCCATTCAGAAACCTTGGCCAT 358
DB 1993 CTTGTCTTGGTGAAGTATGATTTATCTTTGGCCGCCCATTCAGAAACCTTGGCCAT 2052
QY 359 CAAGCCACCAAGAGGCTTCAATTTCTGCTACCTGCTGCTACAGGTTTCCAAACCA 418
DB 2053 CAAGCCACCAAGAGGCTTCAATTTCTGCTACCTGCTGCTACAGGTTTCCAAACCA 2112
QY 419 AAGGCTCAACTGCTGCTACAGCAGGTTA---CTTAGGGCTAAATATTCAGAAAGCACC 474
DB 2113 AAGGCTCAACTGCTGCTACAGCAGGTTA---CTTAGGGCTAAATATTCAGAAAGCACC 2172
QY 475 AAGGCTCAACTGCTGCTACAGCAGGTTA---CTTAGGGCTAAATATTCAGAAAGCACC 501
DB 2173 AAGGCTCAACTGCTGCTACAGCAGGTTA---CTTAGGGCTAAATATTCAGAAAGCACC 2199

RESULT 12
AAS76475
ID AAS76475 standard; cDNA; 4349 BP.
XX
XX AAS76475;
AC
XX
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #12279.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSSE-) HYSSEQ INC.
PA
XX
XX Dmanac RT, Liu C, Tang YT;
PT
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABG12288.
SQ
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1; SEQ ID No 12279; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 4349 BP; 1319 A; 1061 C; 941 G; 1026 T; 2 other;
Query Match 86.8%; Score 435; DB 23; Length 4349;
Best Local Similarity 93.9%; Pred. No. 2.6e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCGAGTATCAATGAGAGCTGTCTCTCATGSCAGAGTATCCTAGCCCT 60
DB 2382 CAAGATCTCGAGTATCAATGAGAGCTGTCTCTCATGSCAGAGTATCCTAGCCCT 2441
QY 61 TATACCTGCTTTCCCAATACAGAGAGAGAGTGTATTAAGTCTGACCTTCAG 120

```
Db 2442 TATACCTGCTTCCCAATACACAGAGAAAGAGTGGTTACAGTCTGATCTTAAAG 2501
Qy 121 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 180
Db 2502 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 2561
Qy 181 ACTTCAACCAACATCTCACTCACTGACTATTTTACCACAGGGTTCCAGGATAGT 240
Db 2562 CCTTCAACCAACATCTCACTCACTGACTATTTTACCACAGGGTTCCAGGATAGT 2621
Qy 241 CCCCATCTATTGGCCAGGCAATTAGCCCAAGACTGAGCCATCTCTACTTGGACACT 300
Db 2622 CCCCATCTATTGGCCAGGCAATTAGCCCAAGACTGAGCCATCTCTACTTGGACACT 2681
Qy 301 --TGCCCTCGGTAGTGGATGATTTACTTTGG--CGCCCAATTCAGAAACCTTGGCCAT 358
Db 2682 CCGTCTCTTCAGTGCATGATGATTTACTTTAG--TGCCCGTTTCAGAAACCTTGGCCAT 2741
Qy 359 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGGCTACATGATTTCCAAACCA 418
Db 2742 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGGCTACATGATTTCCAAACCA 2801
Qy 419 AAGGCTCACTCTGCTCAACAGAGTTA---CTTAGGGCTAAATTAATTCAGAGGCAAC 474
Db 2802 AAGGCTCACTCTGCTCAACAGAGTTAATGCTTAGGGCTAAATTAATTCAGAGGCAAC 2861
Qy 475 AAGGCGCTCAGTGAGAAACATCCAG 501
Db 2862 AAGGCGCTCAGTGAGAAACATCCAG 2888
```

RESULT 13

```
AA576205
ID AA576205 standard; cDNA; 4535 BP.
XX
AC AA576205;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12009.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PE 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG12018.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
PS Claim 1; SEQ ID No 12009; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
```

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful for treating
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AA564197-AA594564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.

XX SQ Sequence 4535 BP; 1356 A; 1159 C; 983 G; 1037 T; 0 other;

Query Match 86.8%; Score 435; DB 23; Length 4535;

Best Local Similarity 93.9%; Pred. No. 2, 7e-130; Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

```
Qy 1 CAAGATCTCAGGATTAATCAATGAGGCTGTTGTTCTGTATAGCCAGCTGTACTAGCCCT 60
Db 1693 CAAGATCTCAGGATTAATCAATGAGGCTGTTGTTCTGTATAGCCAGCTGTACTAGCCCT 1752
Qy 61 TATATCTGCTTTTCCCAATATCCAGAGAAAGAGTGTTCAGTTCAGTTCAGTTCAGTTCAG 120
Db 1753 TATATCTGCTTTTCCCAATATCCAGAGAAAGAGTGTTCAGTTCAGTTCAGTTCAGTTCAG 1812
Qy 121 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 180
Db 1813 GATGCTCTCTCTGATGATCCGTCATCTGATCTCAATCTGTTGGCTTTGAAGAT 1872
Qy 181 ACTTCAACCAACATCTCACTCACTGACTATTTTACCACAGGGTTCCAGGATAGT 240
Db 1873 ACTTCAACCAACATCTCACTCACTGACTATTTTACCACAGGGTTCCAGGATAGT 1932
Qy 241 CCCCATCTATTGGCCAGGCAATTAGCCCAAGACTGAGCCATCTCTACTTGGACACT 300
Db 1933 CCCCATCTATTGGCCAGGCAATTAGCCCAAGACTGAGCCATCTCTACTTGGACACT 1992
Qy 301 --TGCCCTCGGTAGTGGATGATTTACTTTTGGCCCAATTCAGAAACCTTGGCCAT 358
Db 1993 CCGTCTCTTCAGTGCATGATGATTTACTTTTGGCCCAATTCAGAAACCTTGGCCAT 2052
Qy 359 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGGCTACATGATTTCCAAACCA 418
Db 2053 CAAGCCACCCCAAGCGCTCTTCAATTTCTCGCTA--CTGTGGCTACATGATTTCCAAACCA 2112
Qy 419 AAGGCTCACTCTGCTCAACAGAGTTA---CTTAGGGCTAAATTAATTCAGAGGCAAC 474
Db 2113 AAGGCTCACTCTGCTCAACAGAGTTAATGCTTAGGGCTAAATTAATTCAGAGGCAAC 2172
Qy 475 AAGGCGCTCAGTGAGAAACATCCAG 501
Db 2173 AAGGCGCTCAGTGAGAAACATCCAG 2199
```

RESULT 14

```
AA567609
ID AA567609 standard; cDNA; 5154 BP.
XX
AC AA567609;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #3413.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
```



```
QY 132 CTGCATCCCTGTACATGCTGACTCTCAATTCTGTGTTGCGCTTGAAGATACTTCAAAACC 191
    |||
Db 64 CTGCATCCCTGTACATGCTGACTCTCAATTCTGTGTTGCGCTTGAAGATACTTCAAAACC 123
    |||
QY 192 AACATCTCAACTCACTGACTATATTTTACCCCAAGGTTGAGGGATAGTCCCATCTATT 251
    |||
Db 124 AGCATCTCAACTCACTGAGACTATTTTACCCCAAGGTTGAGGGATAGTCCCATCTATT 183
    |||
QY 252 TGGCCAGGCAATTAGCCCAAGACTTGAAGCAATCCCTCATACCTGAGCACTTGTCTTGGT 311
    |||
Db 184 TGGCCAGGCAATTAGCCCAAGACTTGAAGCAATCCCTCATACCTGAGCACTTGTCTTGGT 243
    |||
QY 312 AGGTGATGATTTTCTTTGGCGCGCCCATTCAGAACTTGTGCTGATCAAGCCACCCAG 371
    |||
Db 244 AGGTGATGATTTTCTTTGGCGCGCCCATTCAGAACTTGTGCTGATCAAGCCACCCAG 303
    |||
QY 372 CGCTCTTCAATTTCTGCTGCTACTGTGGCTACATGTTTCCAAACCAAGGCTCACTCT 431
    |||
Db 304 CGCTCTTCAATTTCTGCTGCTACTGTGGCTACATGTTTCCAAACCAAGGCTCACTCT 363
    |||
QY 432 GCTCAGACAGGTTACTTAGGGCTTAAATTAATCCAAAGGCAACGAGGCTTCACTGAGGA 491
    |||
Db 364 GCTCAGACAGGTTACTTAGGGCTTAAATTAATCCAAAGGCAACGAGGCTTCACTGAGGA 423
    |||
QY 492 ACACATCCAG 501
    |||
Db 424 ACACATCCAG 433
    |||
```

Search completed: Apr:19, 2003, 12:18:09
Job time : 210.12 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:03:58 ; Search time 34.6897 Seconds
(without alignments)
4429.120 Million cell updates/sec

Title: US-09-719-554-3_COPY_5500_5500
Perfect score: 501
Sequence: 1 caagatctcagcattatca.....tcagtgcagacacacacag 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/backfillseq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	419.6	83.8	1158	1	US-08-471-724-1
2	419.6	83.8	1158	2	US-08-471-969-1
3	419.6	83.8	1158	2	US-08-384-137-1
4	419.6	83.8	1158	2	US-08-470-006A-1
5	419.6	83.8	1158	3	US-08-691-563C-1
6	419.6	83.8	1158	4	US-09-200-990-1
7	419.6	83.8	1158	4	US-09-133-411-1
8	419.6	83.8	2391	3	US-08-691-563C-57
9	404.4	80.7	1577	3	US-08-691-563C-89
10	271.8	54.3	2330	4	US-09-120-653D-4
11	271.8	54.3	3910	4	US-09-120-653D-1
12	212.6	42.4	297	1	US-08-471-724-2
13	212.6	42.4	297	2	US-08-471-969-2
14	212.6	42.4	297	2	US-08-384-137-2
15	212.6	42.4	297	2	US-08-470-006A-2
16	212.6	42.4	297	3	US-08-691-563C-2
17	212.6	42.4	297	4	US-09-200-990-2
18	212.6	42.4	297	4	US-09-133-411-2
19	211.4	42.2	645	1	US-08-471-724-8
20	211.4	42.2	645	1	US-08-471-969-8
21	211.4	42.2	645	2	US-08-384-137-8
22	211.4	42.2	645	2	US-08-470-006A-8
23	211.4	42.2	645	3	US-08-691-563C-8
24	211.4	42.2	645	4	US-09-200-990-8
25	211.4	42.2	645	4	US-09-133-411-8
26	128.8	25.7	299	3	US-08-691-563C-40
27	116.4	23.2	2448	3	US-08-691-563C-53

28	109.8	21.9	741	1	US-08-471-724-9	Sequence 9, Appl
29	109.8	21.9	741	2	US-08-471-969-9	Sequence 9, Appl
30	109.8	21.9	741	2	US-08-384-137-9	Sequence 9, Appl
31	109.8	21.9	741	2	US-08-470-006A-9	Sequence 9, Appl
32	109.8	21.9	741	3	US-08-691-563C-9	Sequence 9, Appl
33	109.8	21.9	741	4	US-09-200-990-9	Sequence 9, Appl
34	109.8	21.9	741	4	US-09-133-411-9	Sequence 9, Appl
35	109.8	21.9	2389	3	US-08-691-563C-52	Sequence 52, Appl
36	92.6	18.5	4480	4	US-09-167-322-12	Sequence 12, Appl
37	86.6	17.3	6363	2	US-08-929-967-6	Sequence 6, Appl
38	85.8	17.1	1140	4	US-09-603-185-5	Sequence 5, Appl
39	85.8	17.1	3612	4	US-09-265-013-3	Sequence 3, Appl
40	85.8	17.1	7308	4	US-09-011-745-3	Sequence 3, Appl
41	85.8	17.1	7308	4	US-09-011-745-4	Sequence 4, Appl
42	85.8	17.1	7616	4	US-09-011-745-2	Sequence 2, Appl
43	85.8	17.1	8202	1	US-08-258-420-13	Sequence 13, Appl
44	85.8	17.1	8332	3	US-08-850-961-1	Sequence 1, Appl
45	85.8	17.1	8332	4	US-09-479-776-1	Sequence 1, Appl

ALIGNMENTS

```

RESULT 1
; Sequence 1, Application US/08471724
; Patent No. 5800980
; GENERAL INFORMATION:
; APPLICANT: Herve PERRON
; APPLICANT: Francois WALLEY
; APPLICANT: Bernard MANDRAND
; APPLICANT: Frederic BESEME
; TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
; TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOLY
; NUMBER OF INVENTION: 38
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,724
; FILING DATE: June 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 36055C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1158 bases
; TYPE: nucleotide
; STRANDEDNESS: single-stranded
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-471-724-1
;
; Query Match: 83.8%; Score 419.6; DB 1; Length 1158;
; Best Local Similarity: 91.9%; Pred. No. 2.5e-140;
; Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;
;
; QY 1 CAAGATCTCAGATTATCATGAGGCTGTGTTCTCTATAGCAGACTGTACTAGCCCT 60

```

Db 61 CAAGAACTCAGAGTATATCAATGAGGCTGTGTCTCTATACCAAGCTGTATCAACCT 120
Qy 61 TATATCTGCTTTCCCAATATCCAGAGAACAGAGTGTATACAGCTTCAG 120
Db 121 TATACAGCTTTCCCAATATCCAGAGAACAGAGTGTATACAGCTTCAG 180
Qy 121 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTTGAAGAT 180
Db 181 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTTGAAGAT 240
Qy 181 ACTTCAAAACCAATCTCACTCACTGACCTTAACTTCACTTCACTTCACT 240
Db 241 CCTTGAACCAACCTCACTCACTGACCTTAACTTCACTTCACTTCACT 300
Qy 241 CCCCATCTATTTGGCAGGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 298
Db 301 CCCCATCTATTTGGCAGGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 360
Qy 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 361 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGATGATGAT 418
Db 421 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGATGATGAT 480
Qy 419 AAGGCTCACTCTGCTACAGAG---GTTACTTGAAGGCTTAAATTTCAAGGAC 474
Db 481 AAGGCTCACTCTGCTACAGAGATTGATGATGATGATGATGATGATGAT 540
Qy 475 AGGGCCCTCAGTGAAGACATCCAG 501
Db 541 AGGGCCCTCAGTGAAGACATCCAG 567

RESULT 2

US-08-471-969-1, Application US/08471969

Patent No. 5871745

GENERAL INFORMATION:

APPLICANT: Hevea PERRON

APPLICANT: Francois MALLET

APPLICANT: Bernard MANDRAND

APPLICANT: Frederic BESME

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,969

FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 36055A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-471-969-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2,5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

Qy 1 CAAGATTCAGATATCAATGAGGCTGTGTCTCTATACCAAGCTGTATCAACCT 60
Db 61 CAAGAACTCAGAGTATATCAATGAGGCTGTGTCTCTATACCAAGCTGTATCAACCT 120
Qy 61 TATATCTGCTTTCCCAATATCCAGAGAACAGAGTGTATACAGCTTCAG 120
Db 121 TATACAGCTTTCCCAATATCCAGAGAACAGAGTGTATACAGCTTCAG 180
Qy 121 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTTGAAGAT 180
Db 181 GATGCTCTCTCTGATCCCTGATCATCTTCAATCTTGTGCTTTGAAGAT 240
Qy 181 ACTTCAAAACCAATCTCACTCACTGACCTTAACTTCACTTCACTTCACT 240
Db 241 CCTTGAACCAACCTCACTCACTGACCTTAACTTCACTTCACTTCACT 300
Qy 241 CCCCATCTATTTGGCAGGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 298
Db 301 CCCCATCTATTTGGCAGGAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCT 360
Qy 299 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 361 CTGTGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 359 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGATGATGAT 418
Db 421 CAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGCTACATGATGATGAT 480
Qy 419 AAGGCTCACTCTGCTACAGAG---GTTACTTGAAGGCTTAAATTTCAAGGAC 474
Db 481 AAGGCTCACTCTGCTACAGAGATTGATGATGATGATGATGATGATGAT 540
Qy 475 AGGGCCCTCAGTGAAGACATCCAG 501
Db 541 AGGGCCCTCAGTGAAGACATCCAG 567

RESULT 3

US-08-384-137-1

Sequence 1, Application US/08384137

Patent No. 5871996

GENERAL INFORMATION:

APPLICANT: Hevea PERRON

APPLICANT: Francois MALLET

APPLICANT: Bernard MANDRAND

APPLICANT: Frederic BESME

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-384-137-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTTCTGATCCCTGTACATCTGATCTGATCTGATCTGATCTGATCTGAT 180
DB 181 GATGCTCTTCTGATCCCTGTAGCTGATCTGATCTGATCTGATCTGATCTGATCTG 240
QY 181 ACTTCAAAACCAACATCTCACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
DB 241 CCTTGAACCCCAAGCTCTCACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 300
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATCTGATCTGATCTGATCTGATCT 298
DB 301 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATCTGATCTGATCTGATCTGATCT 360
QY 301 CTTGCTCTTCTGATCCCTGTACATCTGATCTGATCTGATCTGATCTGATCTGAT 358
DB 358 CTTGCTCTTCTGATCCCTGTACATCTGATCTGATCTGATCTGATCTGATCTGATCTG 420
QY 358 CAAGCACCCCAAGGCTCTTCAATTTCTGCTACCTGATCTGATCTGATCTGATCTG 418
DB 421 CAAGCACCCCAAGGCTCTTCAATTTCTGCTACCTGATCTGATCTGATCTGATCTG 480
QY 419 AAGGCTCAACTCTGCTCAACAGAG---GTTACTTAAAGGCTTAAATTTATCCAAAGGAC 474
DB 481 AAGGCTCGGCTCTGCTCAACAGAGATTAGATCTAAGGCTTAAATTTATCCAAAGGAC 540
QY 475 AAGGCTCTCACTGAGAGACATCCAG 501
DB 541 AAGGCTCTCACTGAGAGACATCCAG 567

RESULT 4
US-08-470-006A-1
Sequence 1, Application US/08470006A
GENERAL INFORMATION:
APPLICANT: Heine PERRON
APPLICANT: Brandois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BEDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,006A
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-470-006A-1

Query Match 83.8%; Score 419.6; DB 2; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTAGCCCT 60
DB 61 CAAGACTCAGGATTATCAATGAGGCTGTGTTCTCTAATAGCCAGCTGTAGCCCT 120
QY 61 TATACCTGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGGCTTTCCCAATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTTCTGATCCCTGTACATCTGATCTGATCTGATCTGATCTGATCTGAT 180
DB 181 GATGCTCTTCTGATCCCTGTAGCTGATCTGATCTGATCTGATCTGATCTGATCTG 240
QY 181 ACTTCAAAACCAACATCTCACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
DB 241 CCTTGAACCCCAAGCTCTCACTGATCTGATCTGATCTGATCTGATCTGATCTGAT 300
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATCTGATCTGATCTGATCTGATCT 298
DB 301 CCCCATCTATTTGGCCAGGATTAAGCCCAAGATCTGATCTGATCTGATCTGATCT 360
QY 301 CTTGCTCTTCTGATCCCTGTACATCTGATCTGATCTGATCTGATCTGATCTGAT 358
DB 358 CTTGCTCTTCTGATCCCTGTACATCTGATCTGATCTGATCTGATCTGATCTGATCTG 420
QY 358 CAAGCACCCCAAGGCTCTTCAATTTCTGCTACCTGATCTGATCTGATCTGATCTG 418
DB 421 CAAGCACCCCAAGGCTCTTCAATTTCTGCTACCTGATCTGATCTGATCTGATCTG 480
QY 419 AAGGCTCAACTCTGCTCAACAGAG---GTTACTTAAAGGCTTAAATTTATCCAAAGGAC 474
DB 481 AAGGCTCGGCTCTGCTCAACAGAGATTAGATCTAAGGCTTAAATTTATCCAAAGGAC 540
QY 475 AAGGCTCTCACTGAGAGACATCCAG 501
DB 541 AAGGCTCTCACTGAGAGACATCCAG 567

RESULT 5

US-08-691-563C-1

Sequence 1, Application US/08691563C
Patent No. 6001987

GENERAL INFORMATION:

APPLICANT: HERVE PERRON

APPLICANT: FREDERIC BESEME

APPLICANT: FREDERIC BEDIN

APPLICANT: GLAUCIA PARANHOS-BACCALA

APPLICANT: FLORENCE KOMURIAN-PRADEL

APPLICANT: COLETTE JOLIVET

APPLICANT: BERNARD MANDRAND

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 92

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/691,563C

FILING DATE: 02-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 38588

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 base pairs

TYPE: nucleotide

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-691-563C-1

Query Match

Best Local Similarity 91.9%; Pred. No. 2.5e-140;

Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGATTATATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTAGCCCT 60
DB 61 CAAGAACTCAGATTATATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTAGCCCT 120
QY 61 TATACCTGCTTCCCAATATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
DB 121 TATACGAGTCTTCCCAATATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
QY 121 GATGCTTCTTCTGATCCCTGTACATCTGATCTCAATCTTCTGTTGCTTTGAAGAT 180
DB 181 GATGCTTCTTCTGATCCCTGTACATCTGATCTCAATCTTCTGTTGCTTTGAAGAT 240
QY 181 ACTTCAACCCCAATCTCAACTCAGTGTATTTTACCCCAAGGTTTCAAGGAT 240
DB 241 CATTGAACCAACGCTCACTCACTGAGTGTATTACCCCAAGGTTTCAAGGAT 300
QY 241 CCCCATCTATTTGGCAGAGGATTTAGCCCAAGATTTAGCCCAATCTCTATACCTGAGCA-- 298
DB 301 CCCCATCTATTTGGCAGAGGATTTAGCCCAAGATTTAGCCCAATCTCTATACCTGAGCACT 360

RESULT 6

US-09-200-990-1

Sequence 1, Application US/09200990
Patent No. 6184025

GENERAL INFORMATION:

APPLICANT: HERVE PERRON

APPLICANT: FRANCOIS MALLET

APPLICANT: BERNARD MANDRAND

APPLICANT: FREDERIC BEDIN

APPLICANT: FREDERIC BESEME

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INSPECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOLY

TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESSES:

ADDRESSEE: OLIFF & BERRIDGE

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/200,990

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/471,969

FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 36055A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1158 bases

TYPE: nucleotide

STRANDEDNESS: single-stranded

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-200-990-1

QY 299 CTGTGCTTGGAGTGGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGCCAT 358
DB 361 CTGTGCTTGGATGATGATGATTTACTTTGGCCGCCCATTCAGAAACCTTGCCAT 420
QY 359 CAAGCCACCCCAAGCCCTTTCAATTTCTCGTACTGCTGCTACATGTTTCCAAACCA 418
DB 421 CAAGCCACCCCAAGCCCTTTCAATTTCTCGTACTGCTGCTACATGTTTCCAAACCA 480
QY 419 AAGCTCAACTGCTGCTACAGAG---GTTACTAGGGCTAAATTTCCAAAGGACCC 474
DB 481 AAGCTCGGCTCTGCTACAGAGATTGATATCTAGGGCTAAATTTCCAAAGGACCC 540
QY 475 AGGGCCCTCAGTAGAGAACATCCAG 501
DB 541 AGGGCCCTCAGTAGAGAACATCCAG 567

QY 1 CAAGATCTCAGATTATATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTAGCCCT 60
DB 61 CAAGAACTCAGATTATATCAATGAGGCTGTTGTTCTCTATAGCCAGCTGTACTAGCCCT 120
QY 61 TATACCTGCTTCCCAATATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB 121 TATACGAGTCTTCCCAATATACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
QY 121 GATGCTTCTTCTGATCCCTGTACATCTGATCTCAATCTTCTGTTGCTTTGAAGAT 180
DB 181 GATGCTTCTTCTGATCCCTGTACATCTGATCTCAATCTTCTGTTGCTTTGAAGAT 240
QY 181 ACTTCAACCCCAATCTCAACTCAGTGTATTTTACCCCAAGGTTTCAAGGAT 240
DB 241 CATTGAACCAACGCTCACTCACTGAGTGTATTACCCCAAGGTTTCAAGGAT 300
QY 241 CCCCATCTATTTGGCAGAGGATTTAGCCCAAGATTTAGCCCAATCTCTATACCTGAGCA-- 298
DB 301 CCCCATCTATTTGGCAGAGGATTTAGCCCAAGATTTAGCCCAATCTCTATACCTGAGCACT 360

QY 61 TATACCTGCTTTCCCAATACCAAGAGAGAGTGTATTACAGTCTGACCTTACG 120
Db 121 TATACAGTGCTTTCCCAATACCAAGAGAGAGAGTGTATTACAGTCTGACCTTAA 180
QY 121 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGAT 180
Db 181 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGAT 240
QY 181 ACTTCAAAACCAACATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGAT 240
Db 241 CTTTGAACCCCAAGCTCTCACTCACTGATCTGATCTGATCTGATCTGATCTGAT 300
QY 241 CCCCATCTATTGTCGACGAGCATTAAGCCCAAGCTTGAAGCTTCACTGACCA-- 298
Db 301 CCCCATCTATTGTCGACGAGCATTAAGCCCAAGCTTGAAGCTTCACTGACCACT 360
QY 299 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 361 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CAAGCACCACCAAGCGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 418
Db 421 CAAGCACCACCAAGCGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCAACTCTGCTCAACAGCAG---GTTACTTAGGGCTAAATTTATCCAAAGGCA 474
Db 481 AAGGCTCGGCTCTGCTCAACAGGAGATTAATTAAGGGCTAAATTTATCCAAAGGCA 540
QY 475 AAGGCTCTCACTGATGAGAAACATCCAG 501
Db 541 AAGGCTCTCACTGATGAGAAACATCCAG 567

RESULT 7

US-09-133-411-1
Sequence 1, Application US/0913411

Patent No. 6342383

GENERAL INFORMATION:

APPLICANT: Hervé PERRON

APPLICANT: Francois MALLER

APPLICANT: Bernard MANDRAND

APPLICANT: Frederic BESEME

TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR

TITLE OF INVENTION: INSPECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/133,411

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/384,137

FILING DATE: February 6, 1995

ATTORNEY/AGENT INFORMATION:

NAME: Berridge, William P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 36055

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1158 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-133-411-1

Query Match 83.8%; Score 419.6; DB 4; Length 1158;
Best Local Similarity 91.9%; Pred. No. 2.5e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTAATCAATGAGCGTGTGCTCTATTAAGCAAGCTGATCCTGACCT 60
Db 61 CAAGAACTCAGGATTAATCAATGAGCGTGTGCTCTATTAAGCAAGCTGATCCTGACCT 120
QY 61 TATACCTGCTTTCCCAATACCAAGAGAGAGTGTATTACAGTCTGACCTTACG 120
Db 121 TATACAGTGCTTTCCCAATACCAAGAGAGAGTGTATTACAGTCTGACCTTAA 180
QY 121 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGAT 180
Db 181 GATGCTCTTCTGATCCCTGATACCTGATCTGATCTGATCTGATCTGATCTGAT 240
QY 181 ACTTCAAAACCAACATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGAT 240
Db 241 CTTTGAACCCCAAGCTCTCACTCACTGATCTGATCTGATCTGATCTGATCTGAT 300
QY 241 CCCCATCTATTGTCGACGAGCATTAAGCCCAAGCTTGAAGCTTCACTGACCA-- 298
Db 301 CCCCATCTATTGTCGACGAGCATTAAGCCCAAGCTTGAAGCTTCACTGACCACT 360
QY 299 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
Db 361 CTTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 359 CAAGCACCACCAAGCGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 418
Db 421 CAAGCACCACCAAGCGCTCTTCAATTTCTGCTACCTGATGATGATGATGATGATGAT 480
QY 419 AAGGCTCAACTCTGCTCAACAGCAG---GTTACTTAGGGCTAAATTTATCCAAAGGCA 474
Db 481 AAGGCTCGGCTCTGCTCAACAGGAGATTAATTAAGGGCTAAATTTATCCAAAGGCA 540
QY 475 AAGGCTCTCACTGATGAGAAACATCCAG 501
Db 541 AAGGCTCTCACTGATGAGAAACATCCAG 567

RESULT 8

US-08-691-563C-57
Sequence 57, Application US/08691563C

Patent No. 6001987

GENERAL INFORMATION:

APPLICANT: Hervé PERRON

APPLICANT: Frederic BESEME

APPLICANT: Bernard MANDRAND

APPLICANT: Florence KOMORIEN-PRADEL

APPLICANT: Colette JOLIVET

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

NUMBER OF SEQUENCES: 92

CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge

STREET: 700 South Washington Street, Suite 300

CITY: Alexandria

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-57

Query Match 83.8%; Score 419.6; DB 3; Length 2391;
Best Local Similarity 91.9%; Pred. No. 3.8e-140;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 646 CAGGAATCTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 705
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATACAGGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTTTCTTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCT 180
DB 766 GATGCTTTCTTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCT 825
QY 181 ACTTAAACCAACATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCA 240
DB 826 CTTTGAACCAACATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCA 885
QY 241 CCCCATCTATTTGGCCAGGAGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGACA-- 298
DB 886 CCCCATCTATTTGGCCAGGAGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGACA-- 945
QY 299 CTTTGTCTTGGTGTGATGATTTACTTTTGGCCGCCATTTGAGAACTTTTGGCCAT 358
DB 946 CTTTGTCTTGGTGTGATGATTTACTTTTGGCCGCCATTTGAGAACTTTTGGCCAT 1005
QY 359 CAGGACCCAGGAGCTCTCAATTTCTGACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 418
DB 1006 CAGGACCCAGGAGCTCTCAATTTCTGACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1065
QY 419 AAGGCTCACTGTCTCAGAGCAG---GTTACTTAAAGGCTAAATTTATCAAAAGGACAC 474
DB 1066 AAGGCTCGGCTGTCTCAGAGGAGATTAGTACTTAAAGGCTAAATTTATCAAAAGGACAC 1125
QY 475 AGGGCCCTCAGTGTGAGAACATCCAG 501
DB 1126 AGGGCCCTCAGTGTGAGAACATCCAG 1152

RESULT 9
US-08-691-563C-89
Sequence 89, Application US/08691563C
Patent No. 6001987
GENERAL INFORMATION:
APPLICANT: Hervé BERRON
APPLICANT: Frédéric BESME
APPLICANT: Frédéric BÉDIN
APPLICANT: Gaucia PARANHOS-BACCALA

APPLICANT: Florence KOMURIAN-PRADEL
APPLICANT: Colette JOLIVET
APPLICANT: Bernard MANDRAND
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 89:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-691-563C-89

Query Match 80.7%; Score 404.4; DB 3; Length 1577;
Best Local Similarity 89.1%; Pred. No. 8.3e-135;
Matches 449; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGGCTGTACCTAGCCCT 60
DB 998 CAGGAATCTCAGGATTATCAATGAGGCTGTGTCCTCTATAGCCAGGCTGTACCTAGCCCT 1057
QY 61 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 1058 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 121 GATGCTTTCTTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCT 180
DB 1058 TATACCTGCTTTCCCAATATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
QY 121 GATGCTTTCTTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCT 180
DB 1118 GATGCTTTCTTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCTCTGATCT 1177
QY 181 ACTTAAACCAACATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCA 240
DB 1178 CTTTGAACCAACATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCAATCTCA 1237
QY 241 CCCCATCTATTTGGCCAGGAGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGACA-- 300
DB 1238 CCCCATCTATTTGGCCAGGAGATTTAGCCCAAGATTGAGCCAACTCTCATACCTGAGACA-- 1297
QY 301 TGTCTTGTGAGTGTGATGATTTACTTTTGGCCGCCATTTGAGAACTTTTGGCCATCA 360
DB 1238 TGTCTTGTGAGTGTGATGATTTACTTTTGGCCGCCATTTGAGAACTTTTGGCCATCA 1357
QY 361 AGCCACCCAGGAGCTCTCAATTTCTGACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 420
DB 1358 AGCCACCCAGGAGCTCTCAATTTCTGACTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1417
QY 421 GGTCTCACTGTCTCAGAGAGGT---TACTAGGAGCTAAATTTATCAAAAGGACACAG 476
DB 1418 GGTCTCACTGTCTCAGAGAGGT---TACTAGGAGCTAAATTTATCAAAAGGACACAG 1477

QY 477 GGCCCTCAGTGAAGAACATCCA 500
Db 1478 GGCCCTCTGTGAGGAATGATATCCA 1501

RESULT 10

US-09-120-653D-4
; Sequence 4, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-MON
; APPLICANT: JON, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 4
; LENGTH: 2330
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC RETROVIRUS ENDOGENOUS ERV-9
US-09-120-653D-4

Query Match 54.3%; Score 271.8; DB 4; Length 2330;
Best Local Similarity 73.8%; Pred. No. 3.7e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATACCCAGCTGTACTAGCCCT 60
Db 624 CAAGATCTTAACTCACTCAATGACGAGAGTAATCTCTATATCCAGTTGATCCCAACCC 683
QY 61 TATCTCGCTTCCCAATACAGAGAGAGAGTGTGTTACAGTCTGACCTTGAAG 120
Db 684 TATACCTGCTCTCTCAATACAGAGAGAGAGTGTGTTACAGTCTGACCTTGAAG 743
QY 121 GATGCTTCTTCTGATCCCTGTATCATCTGACTCTCAATCTTGTGCTTGAAGAT 180
Db 744 GATGCTTCTTCTGATCCCTGTATCATCTGACTCTGATTCACAGCTCTTGTGCTTGAAGAT 803
QY 181 ACTTCAACCAACATCTCACTGACTGATTTTACCCTAAGGTTCAAGGATAGT 240
Db 804 CCCAACCAACACATCCCACTTACATGATGATGCTTCCCAAGGTTAGGATAGC 863
QY 241 CCCCATCTATTTGGCCAGGATAGCCCAAGACTTGAGCCCAATCTCTCA--TACTGAGCA 298
Db 864 CCTATCTGTTGTGACAGCCCTTACCAAGATCTTCAAGCTTCTGAGTCCAGGCACT 923
QY 299 CTGTGCTTCCGTAGTGAATGATTTACTTTGGCCGCCAATTGAGAAGCTTGGCCAT 358
Db 924 CTGTGCTTCAATATGTGATGATTTACTTTGGCTACAGTTAGAGAGCTTGTGCGAG 983
QY 359 CAAGCCACCAAGGCTTCAATTTCTCGTACCTGTGCTACATGTTTCCAAACCA 418
Db 984 CAGGCTACTTAAATCTCTTGAACCTTCTAGCTAATCAAGGTTACAGAGTGTCTATGTTT 1043
QY 419 AAGGCTCAACTGTGCTACAGAGT----TACTTAAAGGCTTAAATTTCCAAAGGCA 474
Db 1044 AAAGCCCAACTTGTGCTACAGAGGTTAAATATCTTAAAGGCTTAAATCTTAAAGGCA 1103
QY 475 AGGCGCTCAGTGAAGAACATCCAG 501
Db 1104 AGGCGCTCAGGAAGATGAAATACAG 1130

RESULT 11

US-09-120-653D-1
; Sequence 1, Application US/09120653D
; Patent No. 6365727
; GENERAL INFORMATION:
; APPLICANT: YOON, JI-MON
; APPLICANT: JON, HEE-SOOK
; APPLICANT: PARK, HAE-JOON
; APPLICANT: AHN, JONG-SEONG
; APPLICANT: HA, YOUNG-JU
; APPLICANT: CHUNG, SOO-IL
; TITLE OF INVENTION: DNA and peptides of a diabetes-specific endogenous retrovirus
; FILE REFERENCE: 98-338
; CURRENT APPLICATION NUMBER: US/09/120, 653D
; CURRENT FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: KR 98-10108
; PRIOR FILING DATE: 1998-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 1
; LENGTH: 3910
; TYPE: DNA
; ORGANISM: DIABETES-SPECIFIC ENDOGENOUS RETROVIRUS ERV-9
US-09-120-653D-1

Query Match 54.3%; Score 271.8; DB 4; Length 3910;
Best Local Similarity 73.8%; Pred. No. 5e-87;
Matches 374; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTGTTCTCTATACCCAGCTGTACTAGCCCT 60
Db 1020 CAAGATCTTAACTCACTCAATGACGAGAGTAATCTCTATATCCAGTTGATCCCAACCC 1079
QY 61 TATCTCGCTTCCCAATACAGAGAGAGAGTGTGTTACAGTCTGACCTTGAAGAT 120
Db 1080 TATACCTGCTCTCTCAATACAGAGAGAGAGTGTGTTACAGTCTGATTCAGCTTGAAGAT 1139
QY 121 GATGCTTCTTCTGATCCCTGTATCATCTGACTCTCAATCTTGTGCTTGAAGAT 180
Db 1140 GATGCTTCTTCTGATCCCTGTATCATCTGACTCTGATTCACAGCTCTTGTGCTTGAAGAT 1199
QY 181 ACTTCAACCAACATCTCACTGACTGATTTTACCCTAAGGTTCAAGGATAGT 240
Db 1200 CCCAACCAACACATCCCACTTACATGATGATGCTTCCCAAGGTTAGGATAGC 1259
QY 241 CCCCATCTATTTGGCCAGGATAGCCCAAGACTTGAGCCCAATCTCTCA--TACTGAGCA 298
Db 1260 CCTATCTGTTGTGACAGCCCTTACCAAGATCTTAAAGCTTCTGAGTCCAGGCACT 1319
QY 299 CTGTGCTTCCGTAGTGAATGATTTACTTTGGCCGCCAATTGAGAAGCTTGGCCAT 358
Db 1320 CTGTGCTTCAATATGTGATGATTTACTTTGGCTACAGTTAGAGAGCTTGTGCGAG 1379
QY 359 CAAGCCACCAAGGCTTCAATTTCTCGTACCTGTGCTACATGTTTCCAAACCA 418
Db 1380 CAGGCTACTTAAATCTCTTGAACCTTCTAGCTAATCAAGGTTACAGAGTGTCTATGTTT 1439
QY 419 AAGGCTCAACTGTGCTACAGAGT----TACTTAAAGGCTTAAATTTCCAAAGGCA 474
Db 1440 AAAGCCCAACTTGTGCTACAGAGGTTAAATATCTTAAAGGCTTAAATCTTAAAGGCA 1499
QY 475 AGGCGCTCAGTGAAGAACATCCAG 501
Db 1500 AGGCGCTCAGGAAGATGAAATACAG 1526

RESULT 12

US-08-471-724-2
; Sequence 2, Application US/08471724
; Patent No. 580980
; GENERAL INFORMATION:
; APPLICANT: HEVE PERRON
; APPLICANT: FRANCOIS MALLEET
; APPLICANT: BERNARD MANDRAND

APPLICANT: Frederic BREDIN
APPLICANT: Frederic BESEME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
TITLE OF INVENTION: CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,724
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-724-2

Query Match 42.4%; Score 212.6; DB 1; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACTTACCCCT 60
DB 61 CAAGAACTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACTTACCCCT 120
QY 61 TATACCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGTCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTTCTGATCTCCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 180
DB 181 GATGCTCTTCTGATCTCCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
QY 181 ACTTCAACCCCAAGCTCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 235
DB 241 CTTTGAACCCCAAGCTCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 295

RESULT 13
US-08-471-969-2
Sequence 2, Application US/08471969
Patent No. 5871745
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESEME
APPLICANT: Frederic BREDIN
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:

ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,969
FILING DATE: June 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-969-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACTTACCCCT 60
DB 61 CAAGAACTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGTACTTACCCCT 120
QY 61 TATACCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 121 TATACAGTCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 121 GATGCTCTTCTGATCTCCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 180
DB 181 GATGCTCTTCTGATCTCCGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 240
QY 181 ACTTCAACCCCAAGCTCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 235
DB 241 CTTTGAACCCCAAGCTCACTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 295

RESULT 14
US-08-384-137-2
Sequence 2, Application US/08384137
Patent No. 5871996
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MALLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESEME
APPLICANT: Frederic BREDIN
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND BIOPOL
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,137
FILING DATE: February 6, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-384-137-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGAGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTAACCTAGCCCT 60
DB 61 CAAGACTCAGAGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTAACCTAGCCCT 120
QY 61 TATACTGCTTTTCCAAATACAGAGAAAGAGAGTGTTCAGTCTGAGACTTCAG 120
DB 121 TATACAGTGTTCCTCCAAATACAGAGAAAGAGAGTGTTCAGTCTGAGACTTCAG 180
QY 121 GATGCTTTCTTTCGATCCCTGTAACCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
DB 181 GATGCTTTCTTTCGATCCCTGTAACCTGACTCTCAATTTCTGTTGCTTTGAAGAT 240
QY 181 ACTTCAACCCCAACATCTCACTGAGTACTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCCCAACGCTCACTGAGTACTGTTTACCCCAAGGTTCAAGG 295

RESULT 15

US-08-470-006A-2
Sequence 2, Application US/08470006A
Patent No. 5962217
GENERAL INFORMATION:
APPLICANT: Hervé PERRON
APPLICANT: Francois MAILLET
APPLICANT: Bernard MANDRAND
APPLICANT: Frederic BESIME
TITLE OF INVENTION: MSRV1 VIRUS AND MSRV2 PATHOGEN AND/OR
TITLE OF INVENTION: INFECTIVE AGENT ASSOCIATED WITH MULTIPLE SCLEROSIS, AND
TITLE OF INVENTION: BIOPOLYMER CONSTITUENTS THEREOF (AS AMENDED)
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,006A
FILING DATE: June 6, 1995

ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 36055B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 bases
TYPE: nucleotide
STRANDEDNESS: single-stranded
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-470-006A-2

Query Match 42.4%; Score 212.6; DB 2; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.8e-66;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGAGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTAACCTAGCCCT 60
DB 61 CAAGACTCAGAGATTATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTAACCTAGCCCT 120
QY 61 TATACTGCTTTTCCAAATACAGAGAAAGAGAGTGTTCAGTCTGAGACTTCAG 120
DB 121 TATACAGTGTTCCTCCAAATACAGAGAAAGAGAGTGTTCAGTCTGAGACTTCAG 180
QY 121 GATGCTTTCTTTCGATCCCTGTAACCTGACTCTCAATTTCTGTTGCTTTGAAGAT 180
DB 181 GATGCTTTCTTTCGATCCCTGTAACCTGACTCTCAATTTCTGTTGCTTTGAAGAT 240
QY 181 ACTTCAACCCCAACATCTCACTGAGTACTATTTTACCCCAAGGTTCAAGG 235
DB 241 CTTTGAACCCCAACGCTCACTGAGTACTGTTTACCCCAAGGTTCAAGG 295

Search completed: April 19, 2003, 14:48:59
Job time: 38.6897 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 12:08:18 ; Search time 75.717 Seconds
(without alignments)
6646.925 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500
Perfect score: 501
Sequence: 1 caagatctcagatcatca.....tcagtgaagacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 639749 seqs, 502280978 residues

Total number of hits satisfying chosen parameters: 1279498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCITS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	431.8	86.2	1600	7	US-08-979-847-207
2	425.6	85.0	1600	7	US-08-979-847-206
3	422.2	84.3	2304	7	US-08-979-847-87
4	419.6	83.8	1158	7	US-08-979-847-1
5	419.6	83.8	1597	7	US-08-979-847-205
6	419.6	83.8	2365	7	US-08-979-847-88
7	419.6	83.8	2391	7	US-08-979-847-53
8	404.4	80.7	1577	7	US-08-979-847-83
9	402.2	80.3	568	10	US-08-864-761-16537
10	236.6	47.2	557	10	US-08-864-761-11981
11	231.6	46.2	449	10	US-08-864-761-2921
12	221.6	44.2	467	10	US-08-864-761-1194
13	212.6	42.4	297	7	US-08-979-847-2
14	212	42.3	540	10	US-09-864-761-15266
15	211.4	42.2	645	7	US-08-979-847-8
16	182.6	36.4	438	7	US-08-979-847-139
17	182.6	36.4	438	7	US-08-979-847-130
18	182.2	36.4	438	7	US-08-979-847-128
19	182.2	36.4	438	7	US-08-979-847-131

20	181.8	36.3	438	7	US-08-979-847-201	Sequence 201, App
21	165.6	33.1	944	10	US-09-864-761-14911	Sequence 14911, A
22	146.6	29.3	429	7	US-08-979-847-138	Sequence 138, App
23	145.4	29.0	429	7	US-08-979-847-135	Sequence 135, App
24	145	28.9	429	7	US-08-979-847-136	Sequence 136, App
25	145	28.9	429	7	US-08-979-847-137	Sequence 137, App
26	137.2	27.4	150	7	US-08-979-847-36	Sequence 36, App1
27	116.4	23.2	2448	7	US-08-979-847-49	Sequence 49, App1
28	114.2	22.8	191	10	US-09-864-761-119702	Sequence 19702, A
29	109.8	21.9	741	7	US-08-979-847-9	Sequence 9, App1
30	109.8	21.9	2389	7	US-08-979-847-48	Sequence 48, App1
31	107.6	21.5	133	10	US-09-864-761-33921	Sequence 33021, A
32	104.4	20.8	431	10	US-09-864-761-28100	Sequence 28100, A
33	102.2	20.4	181	10	US-09-864-761-17958	Sequence 17958, A
34	92.6	18.5	447	10	US-09-864-761-11512	Sequence 11512, A
35	92.4	18.4	326014	10	US-09-731-231A-3	Sequence 3, App1
36	89	17.8	157	10	US-09-864-761-31788	Sequence 31788, A
37	85.8	17.1	2151	10	US-09-845-157-1	Sequence 1, App1
38	85.8	17.1	8332	10	US-09-006-298-1	Sequence 1, App1
39	82.8	16.5	143068	10	US-09-967-768A-316	Sequence 316, App
40	77.8	15.5	8323	10	US-09-970-597-2	Sequence 2, App1
41	76.8	15.3	183	10	US-09-864-761-28561	Sequence 28561, A
42	76	15.2	685	10	US-09-864-761-31430	Sequence 31430, A
43	70.8	14.1	1436	10	US-09-864-761-17414	Sequence 17414, A
44	69.8	13.9	85	7	US-08-979-847-195	Sequence 195, App
45	69.4	13.9	593	10	US-09-864-761-7065	Sequence 7065, App

ALIGNMENTS

RESULT 1
US-08-979-847-207
Sequence 207, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESIME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMRIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF INVENTIONS: 210
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979, 847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 207:

```

; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1600 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;
US-08-979-847-207

```

Query Match	86.2%;	Score 431.8;	DB 7;	Length 1600;
Best Local Similarity	90.5%;	Pred. No. 5.5e-133;		
Matches 459;	Conservative 20;	Mismatches 22;	Indels 6;	Gaps 2;

Oy	1	CAAGATCTCAGAGTTATCAATGAGGCTGTGTTCCTCTATATACCAGCTGTACCTTACGCTT	60
Db	373	CAGAAGCTCAGAGTTATCAATGAGGCTGTGTTCCTCTATATACCAGCTGTACCTTACGCTT	432
Oy	61	TATACTCTGCTTTCCCAAAATACCAAGAGAAAGCTGAGTGGTTTAACTGCTGGACCTTGAG	120
Db	433	TATATACAGTCTTTCCCAAAATACCAAGAGAAAGCAAGTGGTGTATACAGTCTGAGCCTTAAAG	492
Oy	121	GATGCTCTCTTGATCCCTGTACATCCGTCACTGCAATTTGTATTTGCTTTGAAGAT	180
Db	493	GATGCTCTTTTGTGATCCCTGTACGCTCTGACCTCTCAATTTCTTTGTGCTTTGAAGAT	552
Oy	181	ACTTCAAAACCCAAACATCTCAACTCACTGAGCTATTTTAAACCCAAAGGTTCAAGGATAGT	240
Db	553	CTTTGAACCCCAACGCTCAACCTCACTGAGCTATTTTAAACCCAAAGGTTCAAGGATAGC	612
Oy	241	CCCCATCTATTGGCCAGGCATTAGCCCAAGACTGAGCAATTCCTCACTCACTGGACA--	298
Db	613	CCCCATCTATTGGCCAGGCATTAGCCCAAGACTGAGCAATTTCTCACTCACTGGACAATT	672
Oy	299	CTTGTCTCTCGGTAGGTGATGATTACTTTTGACCGGCCATTCAAGAACTTGTGCCAT	358
Db	673	CTTGTCTCTCGGTAGGTGATGATTACTTTTATGTCRCCTTCAGAAACTTGTGCCAT	732
Oy	359	CAAGCCACCCAGCGCTCTTCAATTCTCGGCTACCTGTGGCTAATAGTTTCCAAAACA	418
Db	733	CAAGCCACCCAGGAACTCTTAATTTCTCTTACCTGTGGCTAATAGGTTTCCAAAACA	792
Oy	419	AAGGCTCAACTCTGCTCAGCAGAGT---TACTTAGGGCTAAATTTATCCAAAGCAC	474
Db	793	AAGGCTCAGCTCTGCTCAGCAGSAGRTTAAATACTTAGGGCTAAATTTATCCAAAGCAC	852
Oy	475	AGGGCCCTTCAGTGAGGAACATCCAG	501
Db	853	AGGCCCTTCAGTGAGGAACATCCAG	879

RESULT 2
 US-08-979-847-206
 , Sequence 206, Application US/08979847
 , Publication No. US20030039664A1
 , GENERAL INFORMATION:
 , APPLICANT: PERRON, HERVE
 , APPLICANT: BESEME, FREDERIC
 , APPLICANT: BEDIN, FREDERIC
 , APPLICANT: PARANHOS-BACCALA, GLAUCIA
 , APPLICANT: KOMURIAN-PRADEL, FLORENCE
 , APPLICANT: JOLIVET-REINAUD, COLETTE
 , APPLICANT: MANDRAND, BERNARD
 , APPLICANT: GARSON, JEREMY
 , APPLICANT: TIKE, PHILIP
 , TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
 , TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
 , TITLE OF INVENTION: THERAPEUTIC PURPOSES;
 , NUMBER OF SEQUENCES: 210
 , CORRESPONDENCE ADDRESS:
 , ADDRESSEE: OLIVF & BERRIDGE, PLC
 , STREET: P. O. BOX 19928
 , CITY: ALEXANDRIA
 , STATE: VA
 , COUNTRY: USA

```

1      ZIP: 22320
2
3      COMPUTER READABLE FORM:
4
5      MEDIUM TYPE: Floppy disk
6
7      COMPUTER: IBM PC compatible
8
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: PatentIn Release #1.0, Version #1.30
11
12     CURRENT APPLICATION DATA:
13
14     APPLICATION NUMBER: US/08/979,847
15
16     FILING DATE: 26-NOV-1997
17
18     CLASSIFICATION: 435
19
20     ATTORNEY/AGENT INFORMATION:
21
22     NAME: BERRIDGE, WILLIAM P.
23     REGISTRATION NUMBER: 30,024
24
25     REFERENCE/DOCKET NUMBER: W2B 39046A
26
27     TELECOMMUNICATION INFORMATION:
28
29     TELEPHONE: 703-836-6400
30     TELEFAX: 703-836-2787
31
32     INFORMATION FOR SEQ ID NO: 206:
33
34     SEQUENCE CHARACTERISTICS:
35
36     LENGTH: 1600 base pairs
37
38     TYPE: nucleic acid
39
40     STRANDEDNESS: single
41
42     TOPOLOGY: linear
43
44     MOLECULE TYPE: DNA (genomic)
45
46     US-08-979-847-206

```

Query Match	85.0%;	Score 425.6;	DB 7;	Length 1600;
Best Local Similarity	88.6%;	Pred. No. 6.3e-131;		
Matches 449;	Conservative 28;	Mismatches 24;	Indels 6;	Gaps 2

QY	1	CAGATCTCAGAGATTATCAAGAGGCGTGTGCTCTCTATAGCCAGCGTACCGAGCCCT	60
Db	373	CAAGAACTCAGAGATTATCAAGAGCGCTGTGTGCTCTATACCAAGCTGACCTAACCT	433
QY	61	TATACTGCTCTTCCCAATATCACAGAGAGACAGAGTGTTACATCTCGACCTTCAG	120
Db	433	TATACAGTGTCTTCCCAATATCCAGAGAGACAGAGTGTTACAGTCTCGACCTTCAG	492
QY	121	GATGCGCTTCTTGTGATCCCTGTACATCTGTACTCTCAATTCCTGTTGCTTTGAAGAT	180
Db	493	GATGCGCTTCTTGTGATCCCTGTACATCTGTACTCTCAATTCCTGTTGCTTTGAAGAT	552
QY	181	ACTTAAACCCCAACATCTCAACTCACTGACGACTATTTTATCCCAAGGTTCAAGGATAGT	240
Db	553	CTTTGAAACCCCAACGCTTCAACTCACTGACGCTTTTATCCCAAGGTTCAAGGATAGC	612
QY	241	CCCATCTATTGTGGCAGGCGATTATGGCCCAACCTTGAGCCAACTCTATACCTGAGACA--	298
Db	613	CCCATCTATTATTGGCGAGGCAATTAGCCCAACCTTGAGCAATCTATACCTGAGACATT	672
QY	299	CTTGTCTTTCGGTAGAGTGATGATTTACTTTTGCGCGGCCATTCAAGAACTTGTGCAT	358
Db	673	CTTGTCTTTCAGTAGATKAGGATGATTTATTMTTATGTCRCCRTTCAGAAACCTTGTGCGAM	733
QY	359	CAAGCCACCCAAAGCGCTCTCAATTTCTCGCTACCTGTGGCTACATGTTTCCAAACCA	418
Db	733	CAAGCCACCCAAAGRHCTCTTAAMTTTCTCTCTACCTGTGGCTACAGGTTCCAAACCA	792
QY	419	AAGGCTCAACTCTGTCTCACAGAGGT----TACTTAGGGCTAAATATATCCAAGGCAAC	477
Db	793	ARGGCTCRSCTCTGTCTCAGASARITTAATATCTNAGGCTAAAAATTATCCAAAGKCRCC	852
QY	475	AGGGGCGCTCAGTGAAGAACATCAAGC	501
Db	853	AGGGGCGCTCAGTGAAGAACATCAAGC	879

RESULT 3
US-08-979-847-87
; Sequence 87, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 87:
SEQUENCE CHARACTERISTICS:
LENGTH: 2304 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-87

Query Match 84.3%; Score 422.2; DB 7; Length 2304;
Best Local Similarity 92.3%; Pred. No. 9.9e-130;
Matches 468; Conservative 0; Mismatches 33; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 640 CAAGATCTCAGGATTATCAATGAGGCTGTTGCTCTATAGCCAGCTGTATCTAGCCCT 699
QY 61 TATACTGCTGCTTCCCAATATCCAGAGAGAGAGGTTTACAGTCTTGACCTTCAAG 120
DB 700 TATACTGCTGCTTCCCTAATACCAAGAGAGAGAGGTTTACAGTCTTGACCTTCAAG 759
QY 121 GATGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 760 GATGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 819
QY 181 ACTTCAAAACCAACATCTCAACTGAGCTATTTTACCCCAAGGGTTCAAGGGATAGT 240
DB 820 CTTTGAACCAACGCTCAACTGAGCTATTTTACCCCAAGGGTTCAAGGGATAGC 879
QY 241 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAGCCATCTCTATCTTGAGCA-- 298
DB 880 CCCCATCTATTTGGCCAGGATTAAGCCCAAGCTTGAGCTATCTCTATCTTGAGCACT 939
QY 299 CTGTCCTTGGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 940 CTGTCCTTGGTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 999
QY 359 CAAGCACCACCAAGCGCTCTCAATTTCTCGCTACCTGTGAGCTAGATGGTTTCAACCA 418

DB 1000 CAAGCACCACCAAGCTCTTACTTCTCTACTGCTGCTCAAGGTTTCCAAACCA 1059
QY 419 AAGGCTCAACTGCTGCTCAAGCAG-----GTTACTAGGGCTAAATTTATCCAAAGCACC 474
DB 1060 AAGGCTGGGCTGCTGCTCAAGCAGGATTAAGTAAAGGCTTAAATTTATCCAAAGCACC 1119
QY 475 AAGGCTGCTCAAGGATTAAGTAAAGGCTTAAATTTATCCAAAGCACC 501
DB 1120 AAGGCTGCTGCTGCTCAAGCAGGATTAAGTAAAGGCTTAAATTTATCCAAAGCACC 1146

RESULT 4
US-08-979-847-1
Sequence 1, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1158 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-1

Query Match 83.8%; Score 419.6; DB 7; Length 1158;
Best Local Similarity 91.9%; Pred. No. 5.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 61 CAAGATCTCAGGATTATCAATGAGGCTGTTGCTCTATAGCCAGCTGTACCTAGCCCT 120
QY 61 TATACTGCTGCTTCCCAATATCCAGAGAGAGAGGTTTACAGTCTTGACCTTCAAG 120
DB 121 TATACAGTCTTCCCAATATCCAGAGAGAGAGGTTTACAGTCTTGACCTTCAAG 180

QY 121 GATGCTTTCTTCGATCCCTGATCCTGATCTGCAATTCCTGTTGCTTTGAAGAT 180
| | | | |
Db 181 GATGCTTTCTTCGATCCCTGATCCTGATCTGCAATTCCTGTTGCTTTGAAGAT 240
| | | | |
QY 181 ACTTCAAAACCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTCCAGGATAGT 240
| | | | |
Db 241 CTTTGAACCCCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTCCAGGATAGT 300
| | | | |
QY 241 CCCCATTATTTGGCAGGATTTAGCCCAAGAC TTGAGCCAAATCCATACCTGAGCA-- 298
| | | | |
Db 301 CCCCATTATTTGGCAGGATTTAGCCCAAGAC TTGAGCCAAATTCCTACCTGAGCACT 360
| | | | |
QY 299 CTTGCTCTCGGTAGGTGATGATTTACTTTTG3CCGCCATTCAGAAAACCTTGCCAT 358
| | | | |
Db 361 CTTGCTCTCGGTAGGTGATGATTTACTTTTA3TCGCCCTTCAGAAAACCTTGCCAT 420
| | | | |
QY 359 CAAGCCACCAAGCGCTTTCAATTTCTCGTACCTGCTGCTAATGTTTCCAAACCA 418
| | | | |
Db 421 CAAGCCACCAAGAACTTTAACTTCTCCTGCTGCTGCTAAGGTTTCCAAACCA 480
| | | | |
QY 419 AAGGCTCAACTGCTGCTCAGCAGCAG---GTTACTTAGGGCTAAATTTATCCAAAGCACC 474
| | | | |
Db 461 AAGGCTCGCTTCTGCTCAGCAGAGATTAGATATCTAAGGCTTAAATTTATCCAAAGCACC 540
| | | | |
QY 475 AAGGCCCTCAGTAGAGAAACATCCAG 501
| | | | |
Db 541 AAGGCCCTCAGTAGAGAAACATCCAG 567
| | | | |

RESULT 5

US-08-979-847-205
; Sequence 205, Application US/08979847
; Publication No. US20030039664A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 205:
SEQUENCE CHARACTERISTICS:
LENGTH: 1597 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-979-847-205

Query Match 83.8%; Score 419.6; DB 7; Length 1597;
Best Local Similarity 91.9%; Pred. No. 6,1e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAAGATCTCAGAGATTATCAATGAGCGTTGTTCTCTATATGCGACGTCATAGCCCT 60
| | | | |
Db 373 CAAGAACTCAGAGATTATCAATGAGCGTTGTTCTCTATATGCGACGTCATAGCCCT 432
| | | | |
QY 61 TATACCTGCTTTCCCAATATCCAGAGAAACAGAGTGTATTAAGTCTGAGACTTCAAG 120
| | | | |
Db 433 TATACAGTCTTTCCCAATATCCAGAGAAACAGAGTGTATTAAGTCTGAGACTTCAAG 492
| | | | |
QY 121 GATGCTTTCTTCGATCCCTGATCCTGATCCTGATCTGCAATTCCTGTTGCTTTGAAGAT 180
| | | | |
Db 493 GATGCTTTCTTCGATCCCTGATCCTGATCCTGATCTGCAATTCCTGTTGCTTTGAAGAT 552
| | | | |
QY 181 ACTTCAAAACCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTCCAGGATAGT 240
| | | | |
Db 553 CTTTGAACCCCAACATCTCAATCTGAGCTATTTTACCCCAAGGTTCCAGGATAGT 612
| | | | |
QY 241 CCCCATTATTTGGCAGGATTTAGCCCAAGAC TTGAGCCAAATCCATACCTGAGCA-- 298
| | | | |
Db 613 CCCCATTATTTGGCAGGATTTAGCCCAAGAC TTGAGCCAAATTCCTACCTGAGCACT 672
| | | | |
QY 299 CTTGCTCTCGGTAGGTGATGATTTACTTTTG3CCGCCATTCAGAAAACCTTGCCAT 358
| | | | |
Db 673 CTTGCTCTCGGTAGGTGATGATTTACTTTTA3TCGCCCTTCAGAAAACCTTGCCAT 732
| | | | |
QY 359 CAAGCCACCAAGCGCTTTCAATTTCTCGTACCTGCTGCTAATGTTTCCAAACCA 418
| | | | |
Db 733 CAAGCCACCAAGAACTTTAACTTCTCCTGCTGCTGCTAAGGTTTCCAAACCA 792
| | | | |
QY 419 AAGGCTCAACTGCTGCTCAGCAGCAG---GTTACTTAGGGCTAAATTTATCCAAAGCACC 474
| | | | |
Db 793 AAGGCTCGCTTCTGCTCAGCAGAGATTAGATATCTAAGGCTTAAATTTATCCAAAGCACC 852
| | | | |
QY 475 AAGGCCCTCAGTAGAGAAACATCCAG 501
| | | | |
Db 853 AAGGCCCTCAGTAGAGAAACATCCAG 879
| | | | |

RESULT 6

US-08-979-847-88
; Sequence 88, Application US/08979847
; Publication No. US20030039664A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 88:
SEQUENCE CHARACTERISTICS:
LENGTH: 2365 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
US-08-979-847-88

Query Match 83.8%; Score 419.6; DB 7; Length 2365;
Best Local Similarity 91.9%; Pred. No. 7.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAATATCTCAGGATTAATCAATGAGGCTGTTCTCTATAGCCAGCTGTAACCTT 60
DB 646 CAAAGACTCAGAGATTAATCAATGAGGCTGTTCTCTATAGCCAGCTGTAACCTT 705
QY 61 TATATCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATATCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTCTTCTGTCATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 180
DB 766 GATGCTCTTCTGTCATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 825
QY 181 ACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCCAAGGTTTCAAGGAT 240
DB 826 CTTTGAACCCCAAGCTCACTCACTGAGCTATTTTACCCCAAGGTTTCAAGGAT 885
QY 241 CCCCATATTTTGGCCAGGATTAAGCCCAAGATTTAGCCCAATCTTACCTGAGCA 298
DB 886 CCCCATATTTTGGCCAGGATTAAGCCCAAGATTTAGCCCAATCTTACCTGAGCA 945
QY 299 CTTGCTCTTGGTATGATGATTTTCTTTGCGCCCAATTTGAGAACTTGTGCCAT 358
DB 946 CTTGCTCTTGGTATGATGATTTTCTTTGCGCCCAATTTGAGAACTTGTGCCAT 1005
QY 359 CAAGCCACCCAGAGGCTCTTCAATTTCTGCTTACCTGCTACATGCTTTCCAAACCA 418
DB 1006 CAAGCCACCCAGAGGCTCTTCAATTTCTGCTTACCTGCTACATGCTTTCCAAACCA 1065
QY 419 AAGGCTCAACTCTGCTCAAGCAG---GTTACTTAGGGCTAAATTTTCCAAAGGACCC 474
DB 1066 AAGGCTCAACTCTGCTCAAGCAG---GTTACTTAGGGCTAAATTTTCCAAAGGACCC 1125
QY 475 AAGGCTCAACTCTGCTCAAGCAG---GTTACTTAGGGCTAAATTTTCCAAAGGACCC 501
DB 1126 AAGGCTCAACTCTGCTCAAGCAG---GTTACTTAGGGCTAAATTTTCCAAAGGACCC 1152

RESULT 7
US-08-979-847-53
; Sequence 53, Application US/08979847
; Publication No. US2003003964A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDIN, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA

APPLICANT: KOMURIAN-BRADEL, FLORENCE
APPLICANT: JOLIVER-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUXE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 2391 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-53

Query Match 83.8%; Score 419.6; DB 7; Length 2391;
Best Local Similarity 91.9%; Pred. No. 7.3e-129;
Matches 466; Conservative 0; Mismatches 35; Indels 6; Gaps 2;

QY 1 CAATATCTCAGGATTAATCAATGAGGCTGTTCTCTATAGCCAGCTGTAACCTT 60
DB 646 CAAAGACTCAGAGATTAATCAATGAGGCTGTTCTCTATAGCCAGCTGTAACCTT 705
QY 61 TATATCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 706 TATATCTGCTTTCCCAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
QY 121 GATGCTCTTCTGTCATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 180
DB 766 GATGCTCTTCTGTCATCTGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 825
QY 181 ACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCCAAGGTTTCAAGGAT 240
DB 826 CTTTGAACCCCAAGCTCACTCACTGAGCTATTTTACCCCAAGGTTTCAAGGAT 885
QY 241 CCCCATATTTTGGCCAGGATTAAGCCCAAGATTTAGCCCAATCTTACCTGAGCA 298
DB 886 CCCCATATTTTGGCCAGGATTAAGCCCAAGATTTAGCCCAATCTTACCTGAGCA 945
QY 299 CTTGCTCTTGGTATGATGATTTTCTTTGCGCCCAATTTGAGAACTTGTGCCAT 358
DB 946 CTTGCTCTTGGTATGATGATTTTCTTTGCGCCCAATTTGAGAACTTGTGCCAT 1005
QY 359 CAAGCCACCCAGAGGCTCTTCAATTTCTGCTTACCTGCTACATGCTTTCCAAACCA 418
DB 1006 CAAGCCACCCAGAGGCTCTTCAATTTCTGCTTACCTGCTACATGCTTTCCAAACCA 1065

Qy 419 AAGCTCACTCTGCTCAACAGAG---GTTACTAGGGCTAAATATTCACAAAGGACCC 474
Db 1066 AAGCTCGCGCTGCTCAACAGAGATTAGATCTAGGCGCTAAATATTCACAAAGGACCC 1125
Qy 475 AGGGCCCTCAGTGAAGAACACATCCAG 501
Db 1126 AGGGCCCTCAGTGAAGAACATTCAG 1152

RESULT 8

US-08-979-847-83
Sequence 83, Application US/08979847
Publication No. US2003003964A1
GENERAL INFORMATION:
APPLICANT: PERRON, HERVE
APPLICANT: BESEME, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TUKE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL ANT. NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
NUMBER OF INVENTIONS: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIEF & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 83:
SEQUENCE CHARACTERISTICS:
LENGTH: 1577 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-83

Query Match 80.7%; Score 404.4; DB 7; Length 1577;
Best Local Similarity 89.1%; Pred. No. 6.7e-124;
Matches 449; Conservative 0; Mismatches 51; Indels 4; Gaps 1;

Qy 1 CAAGATTCAGAGATTATCATGAGGCTGTGTTCCCTCTATAGCAGCTGTACTGAGCCCT 60
Db 998 CAGAGATCTCAGAGATTATATAGGCTGTGTTCCCTCTATAGCAGCTGTACTGAGCCCT 1057
Qy 61 TATATCTGCTTCTCCCAATACCAAGAGAGAGATGTTTACAGTCTTGACCTTCAG 120
Db 1058 TATATCTGCTTCTCCCTATATACCAAGAGAGAGATGTTTACAGTCTTGACCTTCAG 1117
Qy 121 GATGCTCTTCTTGATCCCTGTATACCTGACTC:CAATCTTGTGCTTGAAGAT 180

Db 1118 GATGCTCTTCTTGATCCCTGTATACCTGACTC:CAATCTTGTGCTTGAAGAT 1177
Qy 181 ACTTCAAACCCAAATCTCAATCAGCTGAGTATTTTACCACCAAGGATTCAGGATAGT 240
Db 1178 CTTTGAACCAATGCTCAATTCACCTGAGTCTTTTACCACCAAGGATTCAGGATAGT 1237
Qy 241 CCCCATCTATTGGCCAGGAGTATGACCAAGCTTGAACCAATCTCTATACCTGAGACT 300
Db 1238 CCCCATCTATTGGCCAGGAGTATGACCAAGCTTGAACCAATCTCTATACCTGAGACT 1297
Qy 301 TGTCCTCGTAGGTGATGATTTACTTTTGGCCGACATTCAGAAACCTGTGCAATCA 360
Db 1298 TTGTCCTCGTAGGTGATGATTTTACTTTTGGCCGACATTCAGAAACCTGTGCAATCA 1357
Qy 361 AGCCACCCAAAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGTTTCCAAACCAA 420
Db 1358 AGCCACCCAAAGCGTCTTCAATTTCTCGTACCTGTGCTACATGTTTCCAAACCAA 1417
Qy 421 GCGTCACTCTGCTCAGCAGAGT---TACTTGGGCTAAATTTATCCAAAGGACCCAG 476
Db 1418 GCGTCACTCTGCTCAGCAGAGTAAATTTACTTGGGCTAAATTTATCCAAAGGACCCAG 1477
Qy 477 GGCCCTCAGTGAAGAACATCCCA 500
Db 1478 GGCCCTCAGTGAAGATGTATCCA 1501

RESULT 9

US-09-864-761-16537
Sequence 16537, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 16537
;; LENGTH: 568
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC007567.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.7
US-09-864-761-16537

Query Match 80.3%; Score 402.2; DB 10; Length 568;
Best Local Similarity 91.1%; Pred. No. 2.3e-123;
Matches 462; Conservative 0; Mismatches 38; Indels 7; Gaps 3;

QY 1 CAAGATCTCAGATTATCATAGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 60
DB 58 CAAGATCTCAGATTATCATAGAGGCTGTGTCCTCTATAGCCAGCTGTACCTAGCCCT 117
QY 61 TATACTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 118 TATACTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
QY 121 GATGCTCTCTGATGATCCCTGATCATCTGATCTCAATCTGTTGCTTGAAGAT 180
DB 178 GATGCTCTCTGATGATCCCTGATCATCTGATCTCAATCTGTTGCTTGAAGAT 237
QY 181 ACTTCAACCCCAACATCTCACTCACTGATCTATTTTACCAGAGGTTGAGGATAGT 240
DB 238 CCTTGAACCCCAACATCTCACTCACTGATCTGATCTGATCTGATCTGATCTGATCTG 297
QY 241 CCCCATCTATTGTCGAGGATGAGCCCAAGACTTGAAGCAATCTCTACTGAGCA-- 298
DB 298 CCCCATCTATTGTCGAGGATGAGCCCAAGACTTGAAGCAATCTCTACTGAGCACT 357
QY 299 CTGTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 358
DB 358 CTGTGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 417
QY 359 CAAGCAGCCAGAGGCTCTTCAATTTCTGCTACCTGCTGATGATGATGATGATGATGAT 418
DB 418 CAAGCAGCCAGAGGCTCTTCAATTTCTGCTACCTGCTGATGATGATGATGATGATGAT 477
QY 419 AAGGCTCACTCTGCTCAAGAGG---TTACTTAGGGCTTAAATTTATCCAAAGGACAC 474
DB 478 AA-GCTCAACTCTGCTCAAGAGGCTTAAATTTATCCAAAGGACAC 536
QY 475 AGGGCCCTCAGTAGAGAACATCTCAG 501
DB 537 AGGGCCCTCAGTAGAGAACATCTCAG 563

RESULT 10

US-09-864-761-11981
;; Sequence 11981, Application US/09864761
;; Patent No. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Hanzel, David R.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aesomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; PRIOR FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456

;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 11981
;; LENGTH: 557
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC007923.2
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.96
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
US-09-864-761-11981

Query Match 47.2%; Score 236.6; DB 10; Length 557;
Best Local Similarity 72.2%; Pred. No. 2e-68;
Matches 337; Conservative 0; Mismatches 124; Indels 6; Gaps 2;

QY 41 AGCCAGCTGTAAGCTTATGATCTGCTTTCCAAATACAGAGAGAGAGAGAGAGAGAGT 100
DB 1 ACCCAGTTATACCCACCCCTGATCCCTGCTCTCTCAAAATACAGAGAGAGAGAGATAT 60
QY 101 TTACAGTCTGAGACTTCAGAGATGCTTTCTTGATGATGCTGATGATGATGATGATGAT 160
DB 61 TCAGTGTCTAGAGCTCAAGAGATGCTTTCTTGATGATGCTGATGATGATGATGATGAT 120
QY 161 TCTGTTGACCTTTGAGAGATGCTTCAACCAACCAATCTCACTGATGATGATGATGATGAT 220
DB 121 TTCTCTTTGCTTTGAGAGATGCTTCAACCAACCAATCTCACTGATGATGATGATGATGAT 180
QY 221 CCAGAGGTTCCAGGATAGTCCCATCTATTGTCAGAGAGAGAGAGAGAGAGAGAGAGAG 280
DB 181 CCAGAGGTTTCCAGGATAGTCCCATCTATTGTCAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 281 AATCTCA--TACCTGAGACATGCTGCTTGGTGAAGTGAATGATTTACTTTGGCCGCC 338
DB 241 ACTTCTCAAGTCCAGGACATGCTGCTTGGTGAAGTGAATGATTTACTTTGGCTTCA 300

Qy 339 ATTCGAAACCTTGCCATCAAGCCACCAACGCTCTTCAATTCTCCGCTACTGTG 398
 Db 301 GTTGGAGGCTCATGTGACGAGGCTACTCTAATCTCTTGAATCTTACTATCAAG 360
 Qy 399 GCTCATGTTTCCAAACCAAGGCTCACTCTGCTCAGAGAGT---TACTTAGGGC 454
 Db 361 GGTACAAAGCACTGAGTGGTGAAGGCTCAGCTTACTCTACAGAGGTCAAAATATCTAGGCC 420
 Qy 455 TAAATATTCGAAAGCAGCAGGCGCTCAGTGAAGAACATCCAG 501
 Db 421 TAACTTATCCAGAGGAGCAGGCGCTCAGCAGGAATGATATAG 467
 RESULT 11
 US-09-864-761-2921/c
 Sequence 2921, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aecomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 PRIOR FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263,6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 2921
 LENGTH: 449
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC010856.2
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.6
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
 US-09-864-761-2921
 Query Match 46.2%; Score 231.6; DB 10; Length 449;
 Best Local Similarity 72.8%; Pred. No. 8.4e-67;
 Matches 326; Conservative 0; Mismatches 119; Indels 3; Gaps 2;
 Qy 41 AGCCAGCTGATACCTAGCCCTTATCTGCTTTCCAAATACGAGAGAGAGTGT 100
 Db 449 ATCCAGGCTATACCAACCCCTATACCTGCTCTCAAAATACGAGAGAGAGTGT 390
 Qy 101 TTACAGTCTGAGACCTTCAGAGTGCCTTTCTGATCCCTGTACATCTGACTCAAT 160
 Db 389 TCATTGTTCTGAGACCTCAAGATGCTCTCCCTGCAATTCCTGCACTGACCTCAAG 330
 Qy 161 TCTGTTGCTTGAAGATATCTTCAACCAACATCTCAACCTGACCTGACTATTAC 220
 Db 329 TTTCTTTGCTTTGAGGATCCACAGACACAGCTCCCACTTTCTGATGCTTTGC 270
 Qy 221 CCCAAGGTTGAGGATAGTCCCATCTATTGGCAGCATTAAGCCCAAGCTTGAACC 280
 Db 269 CTCAAGGTTTAGGGATAGCCTTATCTGTTGGTCAAGGCACTGCCCAAGATCTAGGCC 210
 Qy 281 AATCTCA--TACTGACACTTGTCTTGGTAGTGTGATGATTTACTTTGGCCGCC 338
 Db 209 ACTTCTCAAGTACAGGACCTGTGCTTCAAGTAGTGAATTTACTTATGTTTACA 150
 Qy 339 ATTCGAAACCTTGCCATCAAGCCACCAAGGCTCTTCAATTTCTCCGCTACTGTG 398
 Db 149 GTTCAAAAGCTTCATGCGACAGAGCTACTCTAGCTCTCTTGAATCTTACTATCAAG 90
 Qy 399 GCTCATGTTTCCAAACCAAGGCTCACTCTGCTCAC-AGCAGGTTACTTAGGGCTAA 457
 Db 89 GGTACATGGCATCTAAACTGAAGGCCAGCTGCTCAAGTAATATATAGGCTTAA 30
 Qy 458 AATATTCGAAAGCAGCAGGCGCTCTAG 485
 Db 29 TCTTAGCCAGAGAACAGGCGCTCTAG 2
 RESULT 12
 US-09-864-761-1194/c
 Sequence 1194, Application US/09864761
 Patent No. US20020048763A1
 GENERAL INFORMATION:
 APPLICANT: Penn, Sharon G.
 APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 FILE REFERENCE: Aecomica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 PRIOR FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263,6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30


```

PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 1194
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC006999.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.82
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
US-09-864-761-1194

Query Match 44.2%; Score 221.6; DB 10; Length 467;
Best Local Similarity 70.7%; Pred. No. 1.8e-63;
Matches 325; Conservative 0; Mismatches 129; Indels 6; Gaps 2;

```

```

Db 107 GCTGTTAGTCAAAAGCCAGCTTTCCTACAGAGCCAAATATCTAGCTTAATCAT 48
Qy 462 ATCAAGGACCAAGGCGCCCTGAGTGAAGAACATCCAG 501
Db 47 AGCCAGAGGACCAAGGCGCCCTGAGTGAAGAACATCCAG 8

RESULT 13
US-08-979-847-2
Sequence 2, Application US/08979847
Publication No. US20030039664A1
GENERAL INFORMATION:
APPLICANT: PERON, HERVE
APPLICANT: BESEM, FREDERIC
APPLICANT: BEDIN, FREDERIC
APPLICANT: PARANHOS-BACCALA, GLAUCIA
APPLICANT: KOMURIAN-PRADEL, FLORENCE
APPLICANT: JOLIVET-REYNAUD, COLETTE
APPLICANT: MANDRAND, BERNARD
APPLICANT: GARSON, JEREMY
APPLICANT: TURE, PHILIP
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 210
CORRESPONDENCE ADDRESS:
ADDRESSER: OLIVIER & BERRIDGE, PLC
STREET: P.O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-979-847-2

Query Match 42.4%; Score 212.6; DB 7; Length 297;
Best Local Similarity 94.0%; Pred. No. 1.4e-60;
Matches 221; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

Qy 181 ACTTCAACCAACATCTCACTGACATTTTATCCCAAGGATTGAGG 235
Db 241 CCTTTGAACCAACGTCTCACTGACATTTTATCCCAAGGATTGAGG 295

RESULT 14

US-09-864-761-15266/c
; Sequence 15266, Application US/09864761
; Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-09-27

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

Db 540 CCAACCCCTTACCTCTCTCTCAAAATACAGAGAGAGATGTTACAGTTCTG 481
Qy 113 ACCTTGAAGATGCTTCTCTCTGATCCCTGTATACATCTCTCAATTTCTTTGCTT 172
Db 480 AACTCAAGATGCTTCTCTCTGATCCCTGTATACATCTCTCAATTTCTTTGCTT 421

Qy 173 TTGAAGTACTTCAAAACCAACATCTCACTGACATTTATTTTACCCCA-GGGTTC 231
Db 420 TTGAAGTACTTCAAAACCAACATCTCACTGACATTTATTTTACCCCA-GGGTTC 361

Qy 232 AGGATGCTCCATCTATTTGGCAAGGCTTAGCCCAAGTTCAGTCTCTCA-T 289
Db 360 AGGATGCTCCATCTATTTGGCAAGGCTTAGCCCAAGTTCAGTCTCTCA-T 301

Qy 290 ACCTGACATCTCTCTCTGATGATGATTTATTTTGGCCCGCCATTCAGAAAC 349
Db 300 CAGGACATCTCTCTCTGATGATGATTTATTTTGGCCCGCCATTCAGAAAC 241

Qy 350 TTGTCATCAAGCCCAAGGCTCTTCAATTTCTCTGATGATGATTTATTTTGGCC 409
Db 240 TTGTCATCAAGCCCAAGGCTCTTCAATTTCTCTGATGATGATTTATTTTGGCC 181

Qy 410 TCAACCAAGGCTCACTCTGCTCAAGAGT-----TACTTAGGCTAAATATTC 465
Db 180 TCAACCAAGGCTCACTCTGCTCAAGAGT-----TACTTAGGCTAAATATTC 121

Qy 466 AAAGGACCAAGGCTCTGATGATGATTTATTTTGGCCCGCCATTCAGAAAC 501
Db 120 AAAGGACCAAGGCTCTGATGATGATTTATTTTGGCCCGCCATTCAGAAAC 85

RESULT 15
US-08-979-847-8
; Sequence 8, Application US/08979847
; Publication No. US20030039664A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE

APPLICANT: BESEME, FREDERIC

APPLICANT: BEDIN, FREDERIC

APPLICANT: PARANHOS-BACALA, GLAUCIA

APPLICANT: KOMURIAN-PRADEL, FLORENCE

APPLICANT: JOLIVET-REYNAUD, COLETTE

APPLICANT: MANDRAND, BERNARD

APPLICANT: GARSON, JEREMY

APPLICANT: TUKE, PHILIP

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC

TITLE OF INVENTION: THERAPEUTIC PURPOSES

NUMBER OF SEQUENCES: 210

CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIFF & BERRIDGE, PLC

STREET: P.O. BOX 19928

CITY: ALEXANDRIA

STATE: VA

COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847

FILING DATE: 26-NOV-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-8

Query Match 42.2%; Score 211.4; DB 7; Length 645;
Best Local Similarity 89.9%; Pred. No. 5e-60;
Matches 250; Conservative 0; Mismatches 22; Indels 6; Gaps 2;

QY 230 TCAGGATAGTCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAAGCCATCTCAT 289
Db 1 TCAGGATAGCCCCCATCTATTGGCCAGGATTAGCCCAAGACTTGAATTCAT 60
QY 290 ACCTGGACA--CTTGTCTTGGGTAGTGATGATTACTTTGGCCGCCCATTCAGAA 347
Db 61 ACCTGGACACTCTTGTCTTGGGTAGTGATGATTACTTTAGTCGCCGCTCAGAAA 120
QY 348 CCTGTGCCATCAAGGCCCAAGGCTCTCAATTTCTGCTACCTGTGGCTACATG 407
Db 121 CCTGTGCCATCAAGGCCCAAGGCTCTCAATTTCTGCTACCTGTGGCTACATG 180
QY 408 TTTCCAACCAAAAGGCTCACTGTGCTCAGACAG---GTTACTTAGGGCTAAATTAT 463
Db 181 TTTCCAACCAAAAGGCTGGGCTGTGCTCAGAGGATTAAGATTAAGGGCTAAATTAT 240
QY 464 CCAAGGACCAAGGGCCCTCAGTGAAGAAACATCCAG 501
Db 241 CCAAGGACCAAGGGCCCTCAGTGAAGAAACATCCAG 278

Search completed: April 19, 2003, 14:55:23
Job time : 80.717 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2003, 11:45:23 ; Search time 1268.51 Seconds
(without alignments)
6396.429 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatcgcagatcatca.....tcagtgcagacacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_nrc:*
9: gb_estl:*
10: gb_est2:*
11: gb_nrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	384.2	76.7	521	17	AZ517418 RPCI-11-4
2	349	69.7	674	17	AQ111469 CIT-HSP-2
3	324.2	64.7	533	17	B17809 347L9.TVB
4	300	59.9	594	17	AQ193544 CIT-HSP-2
5	266.8	53.0	595	17	B59017 CIT-HSP-201
6	265.6	53.0	745	9	AU121943 AU121943

C 7	256.6	51.2	750	9	AL698185
C 8	253.2	50.5	291	12	BF989696
C 9	252.6	50.4	707	17	B66771
C 10	208	41.5	736	10	AV731083
C 11	206.2	41.2	537	17	AQ270891
C 12	204.4	40.8	416	17	AQ041889
C 13	204.4	40.8	518	17	AQ361037
C 14	203.8	40.7	434	10	BE168410
C 15	201	40.1	408	17	AQ033042
C 16	190	37.9	549	17	AQ231407
C 17	177.8	35.5	557	17	AQ033015
C 18	177.2	35.4	557	17	AQ061550
C 19	175.8	35.1	676	9	AL705035
C 20	168	33.5	518	17	AQ243044
C 21	164.8	32.9	428	17	AQ132081
C 22	161.6	32.3	442	17	AQ075734
C 23	159.2	31.8	541	17	AQ070212
C 24	157.2	31.4	491	17	B17135
C 25	155	30.9	431	17	AQ221075
C 26	152.4	30.4	442	17	AQ0701392
C 27	150.6	30.1	467	17	AQ0876467
C 28	148	29.5	456	17	AQ232635
C 29	148	29.5	519	17	AQ153050
C 30	147.8	29.5	449	17	AQ0701255
C 31	146	29.1	526	17	AQ217989
C 32	143	28.5	332	17	AQ024733
C 33	140.6	28.1	737	17	B67137
C 34	140.4	28.0	437	17	B47870
C 35	140.2	28.0	461	17	AQ225099
C 36	138.4	27.6	518	17	AQ0807540
C 37	132.6	26.5	518	17	AQ284525
C 38	132.2	26.4	429	17	AQ028930
C 39	131.4	26.2	431	17	AQ0736815
C 40	127.2	25.4	433	17	AQ145803
C 41	123	24.6	469	17	AQ223144
C 42	121	24.2	398	17	AQ044697
C 43	119.8	23.9	454	17	AQ226226
C 44	118.2	23.6	533	17	AQ479087
C 45			383	17	AQ192721

ALIGNMENTS

RESULT 1
AZ517418/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

521 bp DNA linear GSS 16-OCT-2000
RPCI-11-4708-TV RPCI-11 Homo sapiens genomic clone RPCI-11-4708,
DNA sequence.
AZ517418
AZ517418.1 GI:10826891
GSS.
human.
Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
,J.C.
BAC end sequences of library RPCI-11
Unpublished (1997)
Other GSSs: RPCI11-4708-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genet cs (info@reagen.com). BAC end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html.
This BAC end was generated during the Red process and may have
higher chance of clone tracking errors.

Seq primer: 17
Class: BAC ends.

FEATURES
Source

Location/Qualifiers
1..521
/organism="Homo sapiens"
/db_xref="GDB:751807"
/db_xref="taxon:9606"
/clone="RPC1-11-4708"
/clone_11b="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.5; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC library"
BASE COUNT 145 a 93 c 147 g 136 t
ORIGIN

Query Match 76.7%; Score 384.2; DB 17; Length 521;

Best Local Similarity 90.8%; Pred. No. 1e-111; Mismatches 38; Indels 6; Gaps 2;
Matches 433; Conservative 0;

31 GTTCTCTATACCAAGCTTACCTTACCTTATCTCTGCTTCCCAATACAGAGAA 90
521 GTTCTCTATACCAAGCTTACCTTATCTCTGCTTCCCAATACAGAGAA 462
91 GCAGAGTGTTCACAGTCTCTGACCTTCAGATCCTTCTTGCATCCCTGATCCT 150
461 GCAGAGTGTTCACAGTCTCTGACCTTCAGATCCTTCTTGCATCCCTGATCCT 402
151 GACTGCAATCTGTTGGCTTGAAGATCTCAACCAACATCTGACCTG 210
401 GACTGCAATCTGTTGGCTTGAAGATCTCTTGAACCAACATCTGACCTG 342
211 ACTATTTTACCCCAAGGTTCAAGGATAGTCCCATCTATTGGCCAGCATTAAGCCAA 270
341 ACTGTTTACCCCAAGGTTCAAGGATAGTCCCATCTATTGGCCAGCATTAAGCCAA 282
271 GACTGAGCAATCTCTGATCTGACA--CTTGTCCTTGGTGGTGGATGATTTACTT 328
281 GACTGAGCAATCTCTGATCTGACA--CTTGTCCTTGGTGGTGGATGATTTACTT 222
329 TTGGCCGCCATTCAGAACCTTGGCCATCAAGTACCCAGCGCTTCAATTTCCCTC 388
221 TTGGCCGCCATTCAGAACCTTGGCCATCAAGTACCCAGCGCTTCAATTTCCCTC 162
389 GCTACCTGTGGCTACATGCTTCCAAACCAAGGTTCAACTCTGCTCAGACAGGT--- 444
161 GCTACCTGTGGCTACATGCTTCCAAACCAAGGTTCAACTCTGCTCAGACAGGT--- 102
445 TACTTAGGGCTAAATTTATTCAGAGGACAGGACCTCAGTGAAGAGACATCCAG 501
101 TACTTAGGGCTAAATTTATTCAGAGGACAGGACCTCAGTGAAGAGATCAG 45

RESULT 2
AQ11469/c 674 bp DNA linear GSS 29-AUG-1998
LOCUS CIT-HSP-2372020.TR CIT-HSP Homo sapiens genomic clone 2372020, DNA
DEFINITION
ACCESSION AQ11469
VERSION AQ11469.1 GI:3488126
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 674)
Adams M.D., Rounsley S.D., Zhao S., Bass S., Linher K., Golden K.,
Berry K., Granger D., Suh E., Wible C., Shizuya H., Simon M. and
Venter J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready

JOURNAL
Map Building
Unpublished (1998)
Other GSSs: CIT-HSP-2372020.TP

COMMENT
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: madams@tigr.org

Clones are available from Research Genetics (info@reagen.com). BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse

FEATURES

Source Location/Qualifiers
1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2372020"
/clone_11b="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 180 a 128 c 184 g 182 t
ORIGIN

Query Match 69.7%; Score 349; DB 17; Length 674;

Best Local Similarity 90.3%; Pred. No. 2.1e-100; Mismatches 30; Indels 14; Gaps 3;
Matches 411; Conservative 0;

53 CTAGCCCTTATACCTGCTTCCCAATACAGAGAGAGAGTTCACGTCGG 112
674 CTAGCCCTTATACCTGCTTCCCAATACAGAGAGAGAGTTCACGTCGG 615
113 ACCTTCAGAGATCCTTCTTGCATCCCTGTACATCTGACTCTCAATCTTGTGCT 172
614 ACCTTCAGAGATCCTTCTTGCATCCCTGTACATCTGACTCTCAATCTTGTGCT 555
173 TTGAAGATCTTCAACCAACATCTCACTGACCTGACATTTTACCCAGGTTCA 232
554 TTGAAGATCTTCAACCAACATCTCACTGACCTGACATTTTACCCAGGTTCA 495
233 GGGATAGTCCCATCTATTGGCCAGGATTAAGCCAGACTGAGCAATCTCATACC 292
494 GGGATAGTCCCATCTATTGGCCAGGATTAAGCCAGACTGAGCAATCTCATACC 435
293 TGGACA--CTTGTCCTTGGTGGTGGATGATTTTACCTTTGGCCGCCATTCAGAACT 350
434 AGGACATCTTGTCTTCAAGTATGATGATTTTACCTTTGGCCGCCATTCAGAACT 375
351 TGTGCGATCAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTT 410
374 TGTGCGATCAAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTT 323
411 CCAACCAAGGCTCAATCTTCTCAGACAGCT---TACTTAGGGCTAAATATCA 466
322 CCAACCAAGGCTCAATCTTCTCAGACAGCT---TACTTAGGGCTAAATATCA 263
467 AAGGACACAGGCTCTGATGAGAGACATCCAG 501
262 AAGGACACAGGCTCTGATGAGAGATGATCCAG 228

RESULT 3

B17809 533 bp DNA linear GSS 04-JUN-1998
LOCUS B17809
DEFINITION 347L9.TVB CIT978SK1 Homo sapiens genomic clone A-347L09, DNA
ACCESSION B17809
VERSION B17809.1 GI:2125558

KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 533)
TITLE	Adams, M.D., Kelley, J.M., Runnaley, S.R. and Venter, J.C.
JOURNAL	Use of a BAC End Sequence Database for Sequence-Ready Map Building Unpublished (1997)
COMMENT	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@igir.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html Seq primer: T7 Class: BAC ends.
FEATURES	Location/Qualifiers
source	1..533 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="A-347109" /clone_11b="CIT978SKA1" /sex="Female" /cell_type="Fibroblast" /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII, Caltech Human BAC Library A1"
BASE COUNT	154 a 97 c 144 g 138 t
ORIGIN	
Query Match	64.7%; Score 324.2; DB 17; Length 533;
Best Local Similarity	86.0%; Pred. No. 1,6e-92;
Matches 398; Conservative	0; Mismatches 53; Indels 12; Gaps 3.
50	TACCTAGCCCTTATCTGCTTGTCCCAATACAGAGGAGCAGAGTGTTACAGTCC 109
533	TACTTAACCCCTATCTGCTTGTCCCAATGACAGAGGAGCAGAGTGTTATAGTCT 474
110	TGGACCTTCAGATGCTTCTGTGATCCCTGTACATCTGAATCTCAATTTGTTTG 169
473	TGGACCTTAAGATGCTTCTTCTGTGATCCGTATCTGACTCTCAATTTGTTTG 414
170	CCTTGAAGATCTTCAACCAACATCTCACTACACCGGACTAATTTACCCCAAGGT 229
413	CTTTGAAGATCTTCTAACCACCAAGTCTCAACTCACCTGGAATGTTTAAACCAAGGT 354
230	TCAGGATAGTCCCATCTATTGGCCAGGACAT-----TAGCCCAAGCTTGAGCAAT 283
353	TCAGGATAGCCCTCATCTATTGCGGAGGACATCAACCAACCAACCTTGAGCACT 294
284	CCTCACTGAGCACT--TGTCCTTGCTAGTGGATGATATTACTTTGGCGCCCAT 341
293	TCTCACTACTGGAACACTCATGTTGTTGGTACATGATGATTTAATCTAGCGGCCCTG 234
342	CAGAAACCTGTGGCATCAAGCACCAAGCCCTCTCAATTCTCCGCTACTGTGGCT 401
233	CAGAAACCTGTGGCATCAAGCACCAAGCTGTTTAACTTCTCCACCATTTGGGCT 174
402	ACATGTTTCCAAACCAAGGCTCAACTGTGCTCAGCAGAGT---TACTTAGGGCTTA 457
173	ACAAGTTTCCAAACCAAGGCTCAGCTGTGCTTATAGCAGGTTAAATCTTAGGGTTAA 114
458	AATTAATCCAAAGCACACAGGCGCTTCAGTGGAGAACACATCCA 500
Db	AATTAATCCAAAGCACACAGGCGCTTCAGGAGGAATGTATCCA 71

LOCUS	AOJ193544	594 bp	DNA	linear	GSS 16-SEP-1998
DEFINITION	CIT-HSP-2381P7.TR CIT-HSP Homo sapiens genomic clone 2381P7, DNA sequence.				
ACCESSION	AOJ193544				
VERSION	AOJ193544.1	GI:3603659			
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 594) Adams,M.D., Rounslev,S.D., Zhao,S., Basa,S., Linbar,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.				
TITLE	Use of a random human BAC End Sequence Database for Sequence-Ready Map Building				
JOURNAL	Unpublished (1998)				
COMMENT	Other GSSs: CIT-HSP-2381P7.TF Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: madams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tcd/humgen/bac_end_search/bac_end_search.html . Seq primer: M13 Reverse Class: BAC ends.				
FEATURES	Location/Qualifiers				
SOURCE	1..594 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="2381P7" /clone_lib="CIT-HSP" /sex="Male" /cell_type="Sperm" /note="Vector: pbeloBAC11; site_1: HindIII; site_2: HindIII"				
BASE COUNT	151 a	111 c	170 g	162 t	
ORIGIN					
Query Match	59.9%; Score 300; DB 17; Length 594;				
Best Local Similarity	91.7%; Pred.No. 9,4e-85;				
Matches 341; Conservative	0; Mismatches 25; Indels 6; Gaps 2;				
OY	136	ATCCCTGACATCTCGACTCTCAATTTCTTTGGCTTTGAAAGATCTTCAAAACCCACA	195		
Db	594	ATCCCTGACATCTCGACTCTCAATTTCTTTGGCTTTGAAAGATCTTCAAACTAATG	535		
OY	196	TCTCAACTCACCTGGACATATTTTACCCCAAGGTTGAGGATAGTCCCATATTTGGC	255		
Db	534	TCTCAACTCACCTGGACATGTTTATACCCCAAGGTTGAGGATAGTCCCATATTTGGC	475		
OY	256	CAGGATTAGCCCAAGACTTGAGCCCAATCCTATCTGAGAC--CTTGCTCTTGATG	313		
Db	474	CAGGACTTAGCCCAAGACTTGAGCCCAATCTATCTGAGACACTTGCTCTTGATG	415		
OY	314	GTGATGATTTACTTTTGGCGCCCATTCAGAAACCTTGTCATCAAGCCCAAGG	373		
Db	414	ATGGATGATTTATTTTGTAGCTGCCCGGTTCAAAACCTTGTCATCAAGCCCAAGG	355		
OY	374	CTCTCAATTTCTGCTACCTGCTGACATGTTTCAAAACCAAGGCTCAACTTGC	433		
Db	354	CTCTTAAATTTCTGCTGCTACCTGCTGACATGTTTCAAAACCAAGGCTCAACTTGC	295		
OY	434	TCACAGCAGGT----TACTTAGGGCTAAATTTATCCAAAGGCACAGGCGCTCAGTGA	489		
Db	294	TCACAGCAGGTTTAATATCTTAGGGCTAAATTTATCCAAAGGCACAGGCGCTCAGTGA	235		
OY	490	GAACATCTCG 501			

Db 234 GAACGTTCCAG 223

RESULT 5
B59017/c 595 bp DNA linear GSS 20-JUN-1998
LOCUS CIT-HSP-2014K19.TR CIT-HSP Homo sapiens genomic clone 2014K19, DNA
DEFINITION sequence.
ACCESSION B59017
VERSION B59017.1 GI:2613735
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 595)
AUTHORS Adams,M.D., Rounley,S.D., Field,C.E., Base,S., Linher,K., Golden
K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
TITLE Unpublished (1997)
JOURNAL Other_GSS: CIT-HSP-2014K19.TF
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
clones are available:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
source
1..595
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:7042821"
/db_xref="taxon:9606"
/clone="2014K19"
/clone_11b="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 158 a 104 c 159 g 174 t
ORIGIN

Query Match 53.3%; Score 266.9; DB 17; Length 595;
Best Local Similarity 89.2%; Pred. No. 3.9e-74;
Matches 323; Conservative 0; Mismatches 32; Indels 7; Gaps 3;

Qy 146 ATCTGACTCTCAATTCCTGTTGCTTTGAAGATACCTTCAAAACCAACATCTCAACTCA 205
Db 595 ATCTGACTCTCAATTCCTGTTGCTTTGAAGATACCTTCAAAACCAACATCTCAACTCA 536
Qy 206 CCTGACTATTTTACCCCAAGGTTCAAGGATAGTCCCATCATTTGGCCAGGCAATTG 265
Db 535 CCTGATGTTTACCCCAAGGTTCAAGGATAGTCCCATCATTTGGCCAGGCAATTG 476
Qy 266 CCCAAGACTTGAACCAATCTCATACCTGAGACATCTGCTTGGGTAGTGATGATT 323
Db 475 CTCAGACTTGAACCAATCTCATACCTGAGACATCTGCTTGGGTAGTGATGATT 416
Qy 324 TACTTTGGCCGCCATTGAGAAACCTTGCCATCAAGCCACCCAGGCTCTTCAATT 383
Db 415 TACTTTTACGCTCCCTTTCAGAAACCTTGCCATCAAGCCACCCAGGATCTTAAATT 356
Qy 384 TCTCTGCTACCTGTGCTACATGCTTCCAAACCAAGGCTCACTCTCTACACAGAG 443
Db 355 TCTCTCCACCTGTGATTAAGGTTTCCAAACCAAGGATCAGCTCTCTACACAGTAG 297

Qy 444 T----TACTTAGGGCTAAATTTATTCAAAGGACCGGCGCTCCTAGTGAGAACATATCC 499
Db 296 TTAATATCTAGTGTAAATTTATTCAAAGGACCGGCGCTCCTAGTGAGAACATATCC 237

Qy 500 AG 501
Db 236 AG 235

RESULT 6
AUI21943 745 bp mRNA linear EST 01-AUG-2002
LOCUS AUI21943 MAMMAL1 Homo sapiens cDNA clone MAMMAL101324 5', mRNA
DEFINITION sequence.
ACCESSION AUI21943
VERSION AUI21943.1 GI:10937178
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 745)
AUTHORS Oca,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
TITLE Unpublished (2000)
JOURNAL Contact: Takao Isogai
COMMENT Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source
1..745
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMAL101324"
/clone_11b="MAMMAL1"
/tissue_type="mammary gland"
/note="Vector: pME185FL3"

BASE COUNT 214 a 196 c 156 g 176 t 3 others
ORIGIN

Query Match 53.0%; Score 265.6; DB 9; Length 745;
Best Local Similarity 89.8%; Pred. No. 1.1e-73;
Matches 309; Conservative 0; Mismatches 29; Indels 6; Gaps 2;

Qy 164 TGTTCGCTTTGAGATACCTTCAAAACCAACATCTCAACTCACTGAGACTATTTTACCC 223
Db 1 TGTTCGCTTTGAGATACCTTCAAAACCAACATCTCAACTCACTGAGACTATTTTACCC 60
Qy 224 AAGGTTGAGGATAGTCCCATCATTTTGGCCAGGCAATTGACCAACTTGAAGCAAT 283
Db 61 AAGGTTGAGGACAGCCCATCATTTTGGCCAGGCAATTGACCAACTTGAAGCAAGT 120
Qy 284 CCTCATACCTGAGACA--CTTGCTCTCGGTAGTGATGATTTACTTTGGCCGCCATT 341
Db 121 TCTCACACTGAGACACTTGTCTCTTCAAGTACATGATGATTTACTTTAGTGCCCTT 180
Qy 342 CAGAAACCTTGTCATGATGAGCAAGCGCTCTTCAATTTCTCTGCTACCTGTGCT 401
Db 181 CAGAAACCTTGTCATGATGAGCAAGCGCTCTTCAATTTCTCTGCTACCTGTGCT 240
Qy 402 ACATGTTTCCAAACCAAGGCTCACTCTGCTACACAGAG----TACTTAGGGCTAA 457
Db 241 ACCAGTTTCCAAACCAAGGCTCACTCTGCTACACAGGCTAAATTAATTAGGGCTAA 300

Qy 458 AATATTCAGGACGAGGCGCTCAGTGAAGAACATCCAG 501
Db 301 AATATTCAGGACGAGGCGCTCAGTGAAGAACATCCAG 344

RESULT 7
AL698185/c 750 bp mRNA linear EST 21-MAR-2002
LOCUS DKFZP686N20106.1 686 (synonym: hlcg3) Homo sapiens cDNA clone
DEFINITION DKFZP686N20106.5, mRNA sequence.
ACCESSION AL698185
VERSION AL698185.1 GI:19618725
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 750)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Well,B. and Wiemann,S.
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Well,B. and Wiemann,S.)
JOURNAL Unpublished (2001)
COMMENT Contact: Ottenwaelder B
MIPS
Am Klopferstr. 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de
Sequenced by MediGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKFZP686N20106) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg; GEMANY; Email: clone@rzpd.de.

FEATURES
source
1..750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP686N20106"
/clone_lib="686 (synonym: hlcg3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SfilA; Site_2: SfilB;
cDNA-collection"

BASE COUNT 195 a 151 c 194 g 210 t

ORIGIN
Query Match 51.2%; Score 256.6; DB 9; Length 750;
Best Local Similarity 73.1%; Pred. No. 8.2e-71;
Matches 372; Conservative 0; Mismatches 129; Indels 8; Gaps 3;

Qy 1 CAAGATCTCAGGATTAATGAGGCTGTTCTCTATAGCAGCTGTACT--AGCC 58
Db 582 CAAGATCTTAAGCTATATGAGGAGATTAATCTCTATATGAGCTGTATCCCAACC 523
Qy 59 CTTAATCTCTGTTTCCCAATATCAGAGAGAGAGAGGTTTACAGTCTCGACCTTC 118
Db 522 CCTAATACCTGCTTCTCAATATCAGAGAGAGAGAGGTTTCACTGTTGAGACCTCA 463
Qy 119 AGGATGCGTCTTCTGCAATCCGTATACCTGACCTCAATCTTGTGCTTTGAAG 178
Db 462 AGGATGCGTCTTCTGCAATCCGTATATTTGAGCTCCAGTTCTCTTTGCTTTGAGG 403
Qy 179 ATACTTCAACCCCAATCTCAACTCAGCTGAGCTATTATCCCAAGGGTTCAAGGATA 238
Db 402 ATCCACAGACCAAGCTCCCAATCTTACATGAGAGCGTCTGCTCAAGGTTTAGGATA 343
Qy 239 GTCCCATCTATTGTCAGGAGGATTAAGCCAGAGCTTGAAGCCAACTCTTA--TACCTGA 296
Db 342 GCCATATCTGTTGGTACGAGGAGCTAGGAGGATCTAGGAGCCACTTCTCAAGTCCAGGCA 283

Qy 297 CACTTGTCCTTCGAGTGTGATGATTTACTTTGGCCGCCATTGAGAACTTTGCGC 356
Db 282 CTGTGTCCTTCAGTATGTGATATTTACTTTGGCTACGAGTTCAAGAGCTTCATGCC 223
Qy 357 ATCAAGCCACCCAGCGCTTTCATTTCCCTGCTACTGTGGCTACATGTTTCCAAAC 416
Db 222 AGCAGGCTACTCTGATGATCTTGAATCTTTAGTAAATCAAGGGTACAGCGCTTAAT 163
Qy 417 CAAGGCTCACTGCTGCTCACAGCAGT---TACTTAAAGGCTAAATATTCAGAGCA 472
Db 162 TGAAGGCCCACTCTGCTTACACAGTCAATATCTAGGCTTAATCTTACCCAGAGGA 103
Qy 473 CCAGGCGCTCAGTGAAGAACATCCAG 501
Db 102 CCAACACCTCAGCAAGATGAAGACAG 74

RESULT 8
BF989696/c 291 bp mRNA linear EST 23-JAN-2001
LOCUS MR2-GN0127-231000-004-e11 GN0127 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF989696
ACCESSION BF989696
VERSION BF989696.1 GI:12396021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 291)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&tl=MR2-GN0127-
231000-004-e11&tl=231000-10-23&tl=1)
Seq primer: puc 18 forward
High quality sequence stop: 290.

FEATURES
source
1..291
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0127"
/dev_stage="Adult"
/note="Organ: placenta, normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORBESTES PCR (U.S. Letters Patent
application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."

BASE COUNT 86 a 54 c 82 g 68 t

ORIGIN
Query Match 50.5%; Score 253.2; DB 12; Length 291;
Best Local Similarity 96.4%; Pred. No. 6.1e-70;
Matches 269; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY	4	GACACAGATATATCAATAGAGCGCTGTGTCTCTATAGCCAGCGCTGACCTAGCCCTAT	63
Db	280	GAACCTTAGGATTAATCAATAGAGCGCTGTGTCTCTATAGCCAGCGCTGACCTAGCCCTAT	221
QY	64	ACTCGCTCTTCCCAATATCCAGAGGAAGCAGAGTGTTTACAGTCTTGACCTTCAGGAT	123
Db	220	ACTCGCTCTTCCCAATATCCAGAGGAAGCAGAGTGTTTACAGTCTTGACCTTCAGGAT	161
QY	124	GCGTTCCTCTGATCCCTGTACATCTGATCTCAATCTGTGTTCGCTTGAATATCT	183
Db	160	GCGTTCCTCTGATCCCTGTACATCTGATCTCAATCTGTGTGTTCGCTTGAATATCT	101
QY	184	TCAACCCCAACATCTCAACTCACTGACTATATTACCCCAAGGTTACAGGATAGTCCC	243
Db	100	TCAACCCCAACATCTCAACTCACTGACTATATTACCCCAAGGTTACAGGATAGTCCC	41
QY	244	CATCTATTTGGCGAGCATATAGCCCAAGACTTGAGCCAA	282
Db	40	CATCTATTTT-GCGAGGCATTTAGCTCTPAGTTAGAGNCCA	3

RESULT 9			
LOCUS	B66771		
DEFINITION	B66771 707 bp DNA linear GSS 21-JUN-1998 CIT-HSP-2015D21.TFB CIT-HSP Homo sapiens genomic clone 2015D21, DNA sequence.		
ACCESSION	B66771		
VERSION	B66771.1	GI:2640749	
KEYWORDS	GSS.		
SOURCE	human.		

ORGANISM	Homo sapiens
EUKARYOTA	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
MAMMALIA	Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 707)
AUTHORS	Adams,M.D., Rounsley,S.D., Field,C.E., Baas,S., Linher,K., Golden, K., Berry,K., Grainger,D., Sun,E., Wible,C., Shizuya,H., Simon,M., and Venter,J.C.
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other GSSs: CIT-HSR-2015D21..TR

Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com), BAC
end search page:
http://www.tigr.org/cdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M3-21
Class: BAC ends.

FEATURES	SOURCE	Location/Qualifiers
		1. .707
		/organism="Homo sapiens"
		/db_xref="GDB:704303g"
		/db_xref="taxon:9606"
		/clone="2015D21"
		/clone_1kb="CTT-HSP"
		/sex="Male"
		/cell_type="Sperm"
		/note=Vector: pBelOAC1, site_1: HindIII, site_2
		HindIII"
BASE COUNT		195 a 183 c 144 g 165 t
ORIGIN		

	Query Match	50.4%	Score 252.6	DB 17	Length 707
	Best Local Similarity	71.4%	Pred. No. 1.5e-69		
	Matches 362	Conservative 0	Mismatches 139	Indels 6	Gaps 2
QY	1 CAAGATCTAGAGATTATCAATGAGCTGTGTTCCTCTATAGCGACCTGTACCTACCCCT	60			

Db	176	CAAGTCTTAGACTCATCAATGAGGCAAGTAAATCTCTAATATCAGGTGATCAACCAACCC	235
Qy	61	TATACTCTGCTTCCCAATAACGAGAGAGAGAGGTTTCAAGTCCGAGACTTCAG	120
Db	236	TATACCTGCCCCCTTCAAAATATACAGAGAAAGCAAGATGTTCACTGTTGGAATTCAG	295
Qy	121	GATGCTCTTCTGTGCATCCCTGTATCTCAGTCTCTCAATCTTGTGTTGGCTTGAAGAT	180
Db	236	GATACCTCTTCTGTCAAGTGTCTTGCACTGTGACTCCCAAGTTTCTTGTGCTTTGAAGAT	355
Qy	181	ACTTCAAACCCACATCTCAACTCACTGGACTATTTTACCACAAGGTTCAAGGATAGT	240
Db	356	CCCCAGAACACCCATCAAACTTACATGACGGCTTGGCTCAAGGTTTAAAGGATAGC	415
Qy	241	CCCCATATTTTGGCCAGGCACTTAGCCCAAGACTTGAAGCAATCTCTA--TACCTGGACA	298
Db	416	CCTCATCCGTTTGGTCAGGACCTGCTGAAGATTTAGGCCACTTCTCAATCCAGGACTT	475
Qy	239	CTTGCTCTCGGTAGTGATGATTTACTTTTGGCCGCCCATTCAGAACTTGTGTCAT	358
Db	476	CTGATGCTTCGGTATGAGATGATTTACTTTTGTCTACAGTTCAGAAAGCCTCATACAG	535
Qy	359	CAAGCACCCAAAGGCTCTTCAATTTCTGTGCTACCTGTGGCTACAGTTCGTTCCAAACA	418
Db	536	CAGGTACTCTAGATCTCTTGAATCTTCTAGCTTATGATGAAGGTTACAAAGGCTCTCAATCA	595
Qy	419	AAGGCTCAACTCTACTACAGCAGAGT----TACTTAGGGCTAAATATTATCAAGGACCC	474
Db	596	AAGGCCAAGCTCTGCTACAAACAAAGTCAATATCTAGGCAATATCTTAGCCAGGAGGCC	655
Qy	475	AGGGCCCTCAGTAGGAACATCCAG	501
Db	656	ATGGCTTTCATCAAGGAAGAAATTCAG	682

RESULT 10					
AV731083/c	AV731083	736 bp	mRNA	linear	EST 17-OCT-2000
LOCUS	AV731083	HTF	Homo sapiens	CDNA clone HTFAH01.5',	mRNA sequence.
DEFINITION	AV731083				
ACCESSION	AV731083				
VERSION	AV731083.1				GI:10840504
KEYWORDS	EST.				

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 736)
AUTHORS	Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, T., Chen, Z. and Han, Z.
TITLE	Homo sapiens cDNA HTF clones

TITLE Homo sapiens cDNA HTF clones
JOURNAL Unpublished (2000)
COMMENT
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source 1. .736
Location/Qualifiers

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTFAAH01"
/clone_id="HTF"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/node=Vector: pbluescript sk(-); Site_1: EcoRI; Site_2
XhoI"
BASE COUNT      193 a      147 c      185 g      211 t

```

ORIGIN

Query Match 41.5%; Score 208; DB 10; Length 736;
Best Local Similarity 74.9%; Pred. No. 2.9e-55;
Matches 274; Conservative 0; Mismatches 90; Indels 2; Gaps 1;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTGTTCTTATAGCCAGCTGTACCTAGCCCT 60
DB 376 CAAGATCTTAGCATCATCATGAGGAGGATTCCTATATACAGTTGTACCAACCCC 317
QY 61 TATATCTGCTTCCCAATATACAGAGAGAGTGTATACAGTCTGTACCTTACG 120
DB 316 TATATCTGCTTCCCAATATACAGAGAGAGTGTATACAGTCTGTACCTTACG 257
QY 121 GATGCTTCTTCTGATCCCTGTATCATCTGATCTCAATCTTGTGCTTGAAGAT 180
DB 256 GATGCTGCTTCTGATCCCTGTATCATCTGATCTCAATCTTGTGCTTGAAGAT 197
QY 181 ACTTCAAACCAACATCTCAATCTGATCTATTTTACCCAGAGGTTCAAGGATAGT 240
DB 196 CCAAGACACACATCAACATCTTGTATGATGTCTTGCCCAAGGTTTATGAGATAGC 137
QY 241 CCCATCTATTGTCGAGGATTTAGCCCAAGCTTGAGCAATCTCA--TACTGAGACA 298
DB 136 CCTCATCTGTTTGTGAGGCACTGGCCCAAGATCTAGCACTTCTTAAAGTCCAGGCACT 77
QY 299 CTGTGCTTGGTAGTGGATTTACTTTTGGCCGCTTCAAGAAACCTTGTGCAAT 358
DB 76 CTGGTCTTCAGTATGATGATATTTACTAATGCTTACCATTTAGAGAAACATGCTCT 17

QY 359 CAAGCC 364
DB 16 CGTGCC 11

RESULT 11 537 bp DNA linear GSS 03-NOV-1998
LOCUS AQ270891/c
DEFINITION H5.2047_A1_H02_MR CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2047 Col=3 Row=O, DNA sequence.
ACCESSION AQ270891
VERSION AQ270891.1 GI:3823431
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 537)
Mahaizra, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.,
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
CONTACT: Mahaizra GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2047 Row: O Column: 3
Class: BAC ends
High quality sequence stop: 537.
Location/Qualifiers
1..537
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2047 Col=3 Row=O"
/clone_1lb="CIT Approved Human Genomic Sperm Library D"
/sex="male"

FEATURES
SOURCE

BASE COUNT 125 a 97 c 146 g 161 t 8 others
ORIGIN

Query Match 41.2%; Score 206.2; DB 17; Length 537;
Best Local Similarity 90.0%; Pred. No. 9.2e-55;
Matches 243; Conservative 0; Mismatches 21; Indels 6; Gaps 2;

QY 238 AGTCCCATCTATTGTCGAGGATTTAGCCCAAGCTTGAACCATCTCATCTGAGAC 297
DB 536 AGCCCCATCTATTGTCGAGGATTTAGCCCAAGCTTGAACCATCTCATCTGAGAC 477
QY 298 A--CTTGTCTTGGTAGTGGATTTACTTTTGGCCGCCATTCAAGAACTTGTGC 355
DB 476 ACTGTGCTTGGTAGTGGATTTACTTTTGGCCGCCATTCAAGAACTTGTGC 417
QY 356 CATCAAGCCACCAAGGCTTCTCAATTTCTGCTACCTGTGCTACATGTTTCCAAA 415
DB 416 CATCAAGCCACCAAGGCTTCTCAATTTCTGCTACCTGTGCTACATGTTTCCAAA 357
QY 416 CCAAGGCTCACTGTGCTACAGAGGTTAAATCTTAAAGGCTTAAATCTTAAAGGCT 471
DB 356 CCAAGGCTCACTGTGCTACAGAGGTTAAATCTTAAAGGCTTAAATCTTAAAGGCT 297
QY 472 ACCAGGCTCTAGTGAAGACATCCAG 501
DB 296 ACCAGACCTTCAGTGAAGACATCCAG 267

RESULT 12 416 bp DNA linear GSS 14-JUL-1998
LOCUS AQ041889/c
DEFINITION CIT-HSP-2335P3.TR CIT-HSP Homo sapiens genomic clone 2335P3, DNA
sequence.
ACCESSION AQ041889
VERSION AQ041889.1 GI:3311083
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.B., Bass, S., Linher, K.,
Simon, M., and Venter, J.C.,
Use of a random BAC end Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
CONTACT: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@igr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.igr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2335P3"
/clone_1lb="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11, Site_1: HindIII, Site_2:
HindIII"

FEATURES
SOURCE

BASE COUNT 119 a 85 c 113 g 99 t

ORIGIN

Seq primer: puc 18 forward
High quality sequence stop: 432.
Location/Qualifiers

FEATURES

source

1..434
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="HT0513"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 125 a 92 c 120 g 97 t
ORIGIN

Query Match 40.7%; Score 203.8; DB 10; Length 434;
Best Local Similarity 73.3%; Pred. No. 4.8e-54;
Matches 288; Conservative 0; Mismatches 102; Indels 3; Gaps 2;

QY 1 CAAGATCTCAGAGTATATATAGAGGCTTGTCTCTATATAGCCAGCTGTACCTAGCCCT 60
DB 413 CAAGATCTTACATCATCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 354
QY 61 TATACCTGCTTCCCAATATACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB 353 G-TACTGTGCTCTCTCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 295
QY 121 GATGCTTCTTCTGATCATCTGATCATCTGATCATCTGATCATCTGATCATCTGATCAT 180
DB 294 GATGCTTCTTCTGATCATCTGATCATCTGATCATCTGATCATCTGATCATCTGATCAT 235
QY 181 ACTTCAACCAACATCTCACTCACTGAGTATTTTACCCCAAGGTTTCAAGGATGT 240
DB 234 CCCAGAGCCACATGTCCCACTTACGTCAGTCTTCCCAAGGCTTACGAGATGC 175
QY 241 CCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 298
DB 174 CTTACCTGTTTGGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 115
QY 299 CTTGCTCTCGTATGATGATGATTTTCTTTGGCCGCGCATTCAGAAACCTTGCCAT 358
DB 114 CTGATCTTCACTATGATGATGATTTTCTTTGGCTACAGTGGGAGGAGGAGGAGGAG 55
QY 359 CAAGCAGCCAGGAGGCTTTCAATTTCTCTGCT 391
DB 54 CAGGCTACTTACATCTCTTGAACCTTCTAGCT 22

RESULT 15 A0033042 408 bp DNA linear GSS 02-JUL-1998
LOCUS HS 2226 A1 D12 MF CIT Approved Human Genomic Sperm Library D Homo

DEFINITION sapiens genomic clone Plate=2226 Col=23 Row=G, DNA sequence.
ACCESSION A0033042
VERSION A0033042.1 GI:3285230
KEYWORDS GSS.

ORGANISM Homo sapiens
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 408)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589
COMMENT High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2226 row: G column: 23
Class: BAC ends
High quality sequence stop: 408.
Location/Qualifiers

FEATURES

source

1..408
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="plate=2226 Col=23 Row=G"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 115 a 83 c 105 g 98 t 7 others
ORIGIN

Query Match 40.1%; Score 201; DB 17; Length 408;
Best Local Similarity 69.5%; Pred. No. 3.7e-53;
Matches 282; Conservative 0; Mismatches 122; Indels 2; Gaps 1;

QY 43 CCAGCTGATACCTTACCTTATCTGCTTCCCAATATACAGAGGAGGAGGAGGAGGAGG 102
DB 408 CCAATGTACCACTATATACATACATCTCTCAATATACAGAGGAGGAGGAGGAGGAGG 349
QY 103 ACACTCTGACCTTACAGATGCTTCTTGCATCCCTGATCCCTGATCCCTGATCCCT 162
DB 348 ACTGTTTGGACNTGAGAGAGGCTTCTTATGATCCATGACCTGCTGCTGCTGCTGCT 289
QY 163 TTGTTGCTTGAAGATCTTGAACCAACATCTCACTCACTGAGTATTTTACC 222
DB 288 CTTTGTCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 229
QY 223 CAAGGTTTACAGGATATCCCATCTATTTGGCCAGGAGGAGGAGGAGGAGGAGGAGG 282
DB 228 CAGGAGTTTACAGGATATCCCATCTATTTGGTCAATATTTGGCCAGGAGGAGGAGG 169
QY 283 TCCTCA--TACCTGAGACCTTGTCTTGGTATGATGATTTTCTTTGGCCGCGCAT 340
DB 168 TTCTCAAGTCCAGGACCTGTGCTCTTCAATATGATGATTTTCTTTGGCTACAGTT 109
QY 341 TCAGAAACCTTGTGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 400
DB 108 TGAAGCTTTCATGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 49
QY 401 TACATGTTTCCAAACCAAGGCTCACTGCTACAGAGGTTA 446
DB 48 TACAAAGATTTAAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3

Search completed: April 19, 2003, 14:46:34
Job time: 1274.51 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 04:28:30 ; Search time 1703.47 Seconds

(without alignments)
8559.314 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500

Perfect score: 501
Sequence: 1 caagatcgcagatcatca.....tcagtgaggaacatccag 501Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: GenBank.*
2: gb_ba.*
3: gb_hcg.*
4: gb_in.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_un.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_hcg_hum.*
39: em_hcg_mus.*
40: em_hcg_other.*
41: em_hcg_hum.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	10499	6 AX007980	AX007980 Sequence
2	501	100.0	56093	6 AX329572	AX329572 Sequence
3	501	100.0	56093	9 HSAC000064	AC000064 Human BAC
4	501	100.0	149194	9 AC007566	AC007566 Homo sapi
5	474	94.6	2938	6 AX000957	AX000957 Sequence
6	474	94.6	2938	6 AX027471	AX027471 Sequence
7	474	94.6	2938	6 AF072499	AF072499 Homo sapi
8	442	88.2	7582	6 AX000966	AX000966 Sequence
9	442	88.2	7582	6 AX027480	AX027480 Sequence
10	435	86.8	40205	6 AF045450	AF045450 Homo sapi
11	435	86.8	142742	9 AF121782	AF121782 Homo sapi
12	435	86.8	340000	9 HS21C080	AL163280 Homo sapi
13	428.6	85.5	8339	9 AL162912	AL162912 Human DNA
14	428.4	85.5	3372	6 AX000964	AX000964 Sequence
15	428.4	85.5	3372	6 AX027478	AX027478 Sequence
16	428.4	85.5	3372	6 AF072502	AF072502 Homo sapi
17	425.4	84.9	146545	2 AC104009	AC104009 Homo sapi
18	423.8	84.6	143063	9 AC040936	AC040936 Homo sapi
19	423.8	84.6	163166	9 AL356632	AL356632 Human DNA
20	423.8	84.6	176773	2 AC009867	AC009867 Homo sapi
21	422.2	84.3	2304	14 AF009668	AF009668 Multiple
22	422.2	84.3	108232	9 AP000654	AP000654 Homo sapi
23	422.2	84.3	156938	9 AL159163	AL159163 Human DNA
24	422.2	84.3	184523	2 AP002890	AP002890 Homo sapi
25	421.2	84.1	139744	9 AL133513	AL133513 Human DNA
26	420.6	84.0	132981	9 HS82711	Z83850 Human DNA s
27	420.6	84.0	161771	9 CNS01DRX	AL121579 Human chr
28	420.6	84.0	178152	2 AC108746	AC108746 Homo sapi
29	420.6	84.0	191426	2 AC026086	AC026086 Homo sapi
30	420.6	84.0	194545	2 AC087482	AC087482 Homo sapi
31	420.6	84.0	194718	2 AC026977	AC026977 Homo sapi
32	420.6	84.0	220807	9 AC093334	AC093334 Homo sapi
33	420.6	84.0	251124	9 HOAB000660	AE000660 Homo sapi
34	419.6	83.8	1158	6 A46012	A46012 Sequence 1
35	419.6	83.8	1158	6 A57048	A57048 Sequence 1
36	419.6	83.8	1158	6 A60113	A60113 Sequence 1
37	419.6	83.8	1158	6 A79463	A79463 Sequence 1
38	419.6	83.8	1158	6 A80197	A80197 Sequence 1
39	419.6	83.8	1158	6 A80306	A80306 Sequence 1
40	419.6	83.8	1158	6 AR035349	AR035349 Sequence
41	419.6	83.8	1158	6 AR036190	AR036190 Sequence
42	419.6	83.8	1158	6 AR036710	AR036710 Sequence
43	419.6	83.8	1158	6 AR076972	AR076972 Sequence
44	419.6	83.8	1158	6 AR094693	AR094693 Sequence
45	419.6	83.8	1158	6 AR183928	AR183928 Sequence

ALIGNMENTS

```
RESULT 1
LOCUS AX007980 10499 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO9967395.
ACCESSION AX007980
VERSION AX007980.1 GI:9995677
KEYWORDS
```

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Perin,J.P., Rieger,F. and Alliel,P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The
library contains cloned DNA from a human male fibroblast cell line
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.
89:8784-8797 (1992); Kim et al., Genomics 34:213-218 (1996).

VECTOR: pBelo

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this
clone is at base position 1 of H_RG083M05; actual end is at 56093
of H_RG083M05

This clone contains STS SW851725.

FEATURES

SOURCE

Location/Qualifiers

1. 56093

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="H_RG083M05"

/clone_1fb="CTB-978SK-B"

/repeat_region

/rpt_family="ALU"

/rpt_family="ALU"

<1360..16971

/gene="WUGSC:H_RG083M05.1"

join(<1360..1503,4181..4370,4587..4774,6422..6556,

9483..9547,11631..11773,11864..12021,13131..13296,

14885..14988,16349..16546,16837..16971)

/gene="WUGSC:H_RG083M05.1"

/note="ATPase; strong similarity to peroxisome

biosynthesis protein Pasi (PID:G1172019); coded for by

human cDNA C04279 (NID:G1465530)"

/codon_start=1

/protein_id="AAB46346.1"

/db_xref="GI:1669371"

/translation="KRLNIQKLEVAFSFAVMNQPSVLLDDLLIAGLPVPEHEH

SPDAORCEILCNVKNKIDCDINKFTDLOHAKKEGVAADFTLVDRATHSRL

SRQISSTRKIVITLDPQKALRGFPLSLSSVNLHKRDKGMKIGLHVRIILMD

TIQLPAKVCLEKKEKPELFLANLPIRQRTGILLYGPFGTKTLGVLARESRKAFISY

KGPBLISKYIGASSQAVADIFIRQAAPCILFDEFEISAPRGHNDVTGVNVNQ

LITQIDGVEGAGVYLAATSRPDLIDPALRPGRLDCVCPDPDQVTTISLESKTG

OMHSFVSRLEINVLSDSLPADVDVLOHVASVTSFTGADIKALINYNLEALHG

MLSKMSRIIPDESKENMYRYPGSSVSESELGNGTSSDLSQCLSAPESTMQDLPV

GKDDLFSGPYLRTASDGCCELTQEQCRDQLRADISITIKGRYRSGEDESMPGPI

KTRLAISQSHLMTALGHTRPSISEDMKNFAEL"

/complement(4948..5130)

/rpt_family="ALU"

/complement(6581..7133)

/rpt_family="L1"

/complement(17767..8037)

/rpt_family="ALU"

/complement(8186..8472)

/rpt_family="ALU"

8473..8625

/gene="WUGSC:H_RG083M05.1"

/note="match to human 3' EST H75782 (NID:G1049794), bases

287-444"

8841..9161

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H75921 (NID:G1050050), bases

21-348"

9481..9547

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST N22627 (NID:G1130501), bases

276-343"

repeat_region

complement(12612..12907)

/rpt_family="ALU"

misc_feature

13670..13793

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:G117434), bases

143-266"

repeat_region

13794..13877

/rpt_family="ALU"

misc_feature

13878..13906

/gene="WUGSC:H_RG083M05.1"

/note="match to human 5' EST H41382 (NID:G117434), bases

30-58"

repeat_region

13907..14104

/rpt_family="ALU"

repeat_region

complement(14110..14137)

/rpt_family="L1"

repeat_region

complement(15618..15907)

/rpt_family="ALU"

repeat_region

17227..17522

/rpt_family="ALU"

misc_feature

18667..19235

/note="match to human fetal brain 5' EST D61494

(NID:G970409), bases 1-255, and to human 3' EST R07476

(NID:G759399)"

repeat_region

19550..19670

/rpt_family="ALU"

misc_feature

21507..37303

/note="similarity to various SS-RNA virus polypeptides;

pseudogene; region of matches and close matches to

multiple human ESTs, see R68740 (NID:G842257)"

37316..37489

/note="Grail prediction, score = 80"

/evidence=not experimental

complement(38938..39224)

/rpt_family="ALU"

39225..39707

/note="match to multiple human ESTs, see N30113

(NID:G1148633)"

39800..40085

/rpt_family="ALU"

repeat_region

complement(40247..40538)

/rpt_family="ALU"

repeat_region

complement(40632..40924)

/rpt_family="ALU"

repeat_region

complement(42283..42891)

/rpt_family="ALU"

repeat_region

complement(44574..45613)

/rpt_family="ALU"

misc_feature

complement(45614..45737)

/note="match to human 3' EST H48898 (NID:G988738), bases

129-333"

complement(46107..47026)

/note="match to multiple human ESTs, see N81064

(NID:G1243765), H48897 (NID:G988737), and W78831

(NID:G273146)"

complement(47027..47318)

/rpt_family="ALU"

repeat_region

complement(47365..47782)

/note="match to multiple human ESTs, see W37495

(NID:G1319089)"

47898..48115

/note="match to human 5' EST H62306 (NID:G1015138), bases

93-368"

complement(48116..48405)

/rpt_family="ALU"

misc_feature

complement(48406..48584)

/note="match to human 3' EST N29952 (NID:G1148472), bases

290-455, and 5' EST R12730 (NID:G765806)"

complement(48787..49405)

/rpt_family="ALU"

misc_feature

complement(49406..49534)

/note="match to human 3' EST R65794 (NID:G838432), bases

309-440"

```

repeat_region      complement(49638..45672)
                    /rpt_family="ALU"
misc_feature        /note="match to human 3', EST N29952 (NID:g1148472) and 5'
                    EST N29938 (NID:g1148458), sequences are from opposite
                    ends of the same clone"
gene                /gene="WUGSC:H.RG083M05.2"
                    /complement(49698..51806)
                    /note="WUGSC:H.RG083M05.2"
                    /complement(join(49698..49888,51575..51806))
                    /note="coded for by human cDNA# W37389 (NID:g1319205),
                    R65891 (NID:g838529), R65794 (NID:g838432) and R65794
                    (NID:g838432)"
CDS                 /codon_start=1
                    /protein_id="AAB46345.1"
                    /db_xref="GI:1669370"
                    /translation="MFYFQVGIIFPCPGVYVIOGVVSVIEBQDKPYAAIRGR
                    IQDYCEKSAALWLIPTLSIRDFDPSYIIGEEEDLPKMEYLEFVCHAPSEYFK
                    SRSSFPFVPRPEKGYIWMVGPPIPAITIKESVANHL"
exon                 /complement(51576..51758)
                    /gene="WUGSC:H.RG083M05.2"
                    /note="WUGSC:H.RG083M05.2"
                    /note="Grail prediction, score = 86"
                    /evidence=not_experimental
repeat_region       /complement(52052..52329)
                    /rpt_family="11"
misc_feature         /note="match to human EST W79192 (NID:g273505) base 2-289"

Query Match        100.0%; Score 501; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 2,36-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

ACCESSION          AC007566
VERSION            AC007566.2
KEYWORDS           GI:11181861
SOURCE             HTG.
ORGANISM           Homo sapiens.
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE          1 (bases 1 to 149194)
AUTHORS           Sulston, J.E. and Waterston, R.
TITLE             Toward a complete human genome sequence
JOURNAL           Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE           99063792
PUBMED            9847074
REFERENCE          2 (bases 1 to 149194)
AUTHORS           Du, Z.
TITLE             The sequence of Homo sapiens BAC clone CTB-10G5
JOURNAL           Unpublished (2001)
REFERENCE          3 (bases 1 to 149194)
AUTHORS           Waterston, R.H.
TITLE             Direct Submission
JOURNAL           Submitted (15-MAY-1999) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE          4 (bases 1 to 149194)
AUTHORS           Waterston, R.H.
TITLE             Direct Submission
JOURNAL           Submitted (02-OCT-2000) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE          5 (bases 1 to 149194)
AUTHORS           Waterston, R.H.
TITLE             Direct Submission
JOURNAL           Submitted (16-NOV-2000) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE          6 (bases 1 to 149194)
AUTHORS           Waterston, R.H.
TITLE             Direct Submission
JOURNAL           Submitted (03-JAN-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE          7 (bases 1 to 149194)
AUTHORS           Waterston, R.H.
TITLE             Direct Submission
JOURNAL           Submitted (06-FEB-2002) Genome Sequencing Center, Washington
                   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                   MO 63108, USA
REFERENCE          8 (bases 1 to 149194)
AUTHORS           Waterston, R.
TITLE             Direct Submission
JOURNAL           Submitted (01-MAR-2002) Department of Genetics, Washington
                   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
                   On Nov 16, 2000 this sequence version replaced gi:4835815.
COMMENT            ----- Genome Center
                   Center: Washington University Genome Sequencing Center
                   Center code: WUGSC
                   Web site: http://genome.wustl.edu/gsc
                   Contact: sapiens@watson.wustl.edu
                   ----- Summary Statistics
                   -----
                   Center project name: H_RG010605

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-9785K-B. The library contains cloned DNA from the male fibroblast cell line 9785K. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelOBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

FEATURES

source

Location/Qualifiers
1..149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_lib="CTB-9785K-B"
1..1634
/rpt_family="L1"
repeat_region
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
repeat_region
2696..3066
/rpt_family="MaLR"
3108..3392
/rpt_family="Alu"
3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
3542..3628
/note="match to EST BG766882 (NID:g14077535)"

3542..3628
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
3542..3609
/note="match to EST AW579261 (NID:g7254310)"
misc_feature
3562..3628
/note="match to EST BE272564 (NID:g91834561)"
misc_feature
3598..3628
/note="match to EST BE299708 (NID:g91834561)"
misc_feature
3714..3785
/note="match to EST BG260659 (NID:g12770475)"
misc_feature
3717..3785
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"
misc_feature
3719..3785
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature
3719..3785
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
3719..3785
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
3719..3785
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
3719..3785
/note="match to EST BE299708 (NID:g91834561)"
misc_feature
3719..3785
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
3719..3785
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
3719..3785
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature
3860..3862
/note="match to EST AV686676 (NID:g10288539)"
repeat_region
4730..4827
/rpt_family="L2"
4834..4925
/rpt_family="L1"
misc_feature
5715..5860
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
5715..5860
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
5715..5860
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
5715..5860
/note="match to EST BE299708 (NID:g91834561)"
misc_feature
5715..5860
/note="match to EST BG260659 (NID:g12770475)"
misc_feature
5715..5860
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
5715..5860
/note="match to EST BI160365 (NID:g14620366)"
misc_feature
5715..5860
/note="similar to Mus musculus EST BE994936 (NID:g10678674)"

Query Match 100.0%; Score 501; DB 9; Length 149194;
Best Local Similarity 100.0%; Pred. No. 2, 2e-138;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTAATCATGAGGCTGTCTCTATAGCCAGCTGTACTAGCCCT 60
DB 89004 CAAGATCTCAGGATTAATCAAGAGGCTGTCTCTATAGCCAGCTGTACTAGCCCT 88945
QY 61 TATATCTGCTTCCCAATATCCAGAGGAGCAGAGTGTATACAGTCTGACCTTCAG 120
DB 88944 TATATCTGCTTCCCAATATCCAGAGGAGCAGAGTGTATACAGTCTGACCTTCAG 88885
QY 121 GATGCTTCTTCTGATCCCTGTACATCTGACTCTCAATCTTGTGCTTGAAGAT 180
DB 88884 GATGCTTCTTCTGATCCCTGTACATCTGACTCTCAATCTTGTGCTTGAAGAT 88825
QY 181 ACTTCAACCCCAATCTCACTCACTGTGACTATTTTACCCCAAGGTTTACGGATAGT 240
DB 88824 ACTTCAACCCCAATCTCACTCACTGTGACTATTTTACCCCAAGGTTTACGGATAGT 88765

QY 241 CCCCATCTATTGGCCAGGCAATTAGCCCAAGATTGAGCCCAATCTCTATCTGACACT 300
Db 88764 CCCCCTATTATTTGGCCAGGCAATTAGCCCAAGATTGAGCCCAATCTCTATCTGACACT 88705
QY 301 TGTCTTGGTATGATGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGGCA 360
Db 88704 TGTCTTGGTATGATGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGGCA 88645
QY 361 AGCCACCCAGGCGCTCTCAATTTCTGCTACCTGTGAGTATGATGTTTCCAAACCAA 420
Db 88644 AGCCACCCAGGCGCTCTCAATTTCTGCTACCTGTGAGTATGATGTTTCCAAACCAA 88585
QY 421 GGCTCACTGTGCTCAGCAGAGGTTACTTACGCTAAATATTCAAAGGCAACGAGCC 480
Db 88584 GGCTCACTGTGCTCAGCAGAGGTTACTTACGCTAAATATTCAAAGGCAACGAGCC 88525
QY 481 CTGAGTGAGAACATCTCAG 501
Db 88524 CTGAGTGAGAACATCTCAG 88504

RESULT 5
AX000957 2938 bp DNA linear PAT 10-MAR-2000
LOCUS AX000957
DEFINITION Sequence 2 from Patent WO902695.
ACCESSION AX000957
VERSION AX000957.1 GI:7241199
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 2938)
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 2 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source 1..2938
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAAGATCTCAGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACCTGAGCC 59
Db 2079 CAAGATCTCAGATTATCAATGAGGCGGTGTCTTTATACCAGCTGACCTGAGCC 2138
QY 60 TTAATCTGTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
Db 2139 TTAATCTGTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
QY 120 GGAATGCTTCTTGCATCCCTGTACATCTGACCTGACCTGACCTGACCTGACCTGAC 179
Db 2199 GGAATGCTTCTTGCATCCCTGTACATCTGACCTGACCTGACCTGACCTGACCTGAC 2258
QY 180 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTGAGGATAG 239
Db 2259 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTGAGGATAG 2318
QY 240 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCCAATCTCTATCTGAGAC 299
Db 2319 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCCAATCTCTATCTGAGAC 2378
QY 300 TGTCTTGGTATGATGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGGCA 359
Db 2379 TGTCTTGGTATGATGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGGCA 2438
QY 360 AAGCCACCAAGCGCTCTCAATTTCTGCTACCTGTGAGTATGATGTTTCCAAACCAA 419

Db 2439 AAGCCACCAAGCGCTCTCAATTTCTGCTACCTGTGAGTATGATGTTTCCAAACCAA 2498
QY 420 AGCTCAACTGTGCTCAGAGAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGAGCC 479
Db 2499 AGCTCAACTGTGCTCAGAGAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGAGCC 2558
QY 480 CTTGAGTGAGAACATCTCAG 501
Db 2559 CTTGAGTGAGAACATCTCAG 2580

RESULT 6
AX027471 2938 bp DNA linear PAT 16-SEP-2000
LOCUS AX027471
DEFINITION Sequence 21 from Patent FR278784.
ACCESSION AX027471
VERSION AX027471.1 GI:10188435
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2938)
AUTHORS Mallet, F., Voisset, C. and Paranhos, B.G.
JOURNAL Patent: FR 278784-A 21 28-JUL-2000;
BIO MERIEUX (FR)
FEATURES
source 1..2938
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN

Query Match 94.6%; Score 474; DB 6; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 CAAGATCTCAGATTATCAAT-GAGGCTGTGTTCTCTATAGCCAGCTGACCTGAGCC 59
Db 2079 CAAGATCTCAGATTATCAATGAGGCGGTGTCTTTATACCAGCTGACCTGAGCC 2138
QY 60 TTAATCTGTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
Db 2139 TTAATCTGTGCTTTCCCAATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
QY 120 GGAATGCTTCTTGCATCCCTGTACATCTGACCTGACCTGACCTGACCTGACCTGAC 179
Db 2199 GGAATGCTTCTTGCATCCCTGTACATCTGACCTGACCTGACCTGACCTGACCTGAC 2258
QY 180 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTGAGGATAG 239
Db 2259 TACTTCAACCAACATCTCACTCACTGACTATTTTACCCCAAGGTTGAGGATAG 2318
QY 240 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCCAATCTCTATCTGAGAC 299
Db 2319 TCCCATCTATTGGCCAGGCAATTAGCCCAAGACTTGAGCCCAATCTCTATCTGAGAC 2378
QY 300 TGTCTTGGTATGATGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGGCA 359
Db 2379 TGTCTTGGTATGATGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGGCA 2438
QY 360 AAGCCACCAAGCGCTCTCAATTTCTGCTACCTGTGAGTATGATGTTTCCAAACCAA 419
Db 2439 AAGCCACCAAGCGCTCTCAATTTCTGCTACCTGTGAGTATGATGTTTCCAAACCAA 2498
QY 420 AGCTCAACTGTGCTCAGAGAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGAGCC 479
Db 2499 AGCTCAACTGTGCTCAGAGAGGTTACTTAAAGGCTAAATTTATCCAAAGCAGGAGCC 2558
QY 480 CTTGAGTGAGAACATCTCAG 501
Db 2559 CTTGAGTGAGAACATCTCAG 2580

RESULT 7
LOCUS AF072499 2938 bp mRNA linear PRI 10-FEB-1999
DEFINITION Homo sapiens endogenous retrovirus W sequence.
ACCESSION AF072499
VERSION AF072499.1 GI:4262283
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2938)
AUTHORS Blond, J.L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H.,
Mandrand, B. and Mallet, F.
TITLE Molecular characterization and placental expression of HERV-W, a
new human endogenous retrovirus family
JOURNAL J. Virol. 73 (2), 1175-1185 (1999)
MEDLINE 9882319
PUBMED 9882319
REFERENCE 2 (bases 1 to 2938)
AUTHORS Blond, J.L., Beseme, F. and Mallet, F.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1998) UMI03 CNRS-Biomelex, ENS Lyon, 46 allée
d'Italie, Lyon, Cedex 07 69364, France
FEATURES
source location/Qualifiers
1..2938 /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cl.6A1"
/cissue_type="placenta"
/clone_lib="Clontech 5'-stretch plus library, Cat number
HL50143"
misc_feature 1..2938
/note="corresponds to gag and pol"
repeat_region 1..2938
/note="HERV-W"
BASE COUNT 878 a 720 c 646 g 692 t 2 others
ORIGIN
Query Match 94.6%; Score 474; DB 9; Length 2938;
Best Local Similarity 97.8%; Pred. No. 3e-130;
Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 1 CAAGATCTCAGATTAAT-GAGGCTGTGTCCTCTATAGCCAGCTGTAAGCC 59
DB 2079 CAAGATCTCAGATTAATCAATGAGGCGTGTCTTTATACCAAGCTGTAAGCC 2138
QY 60 TTATCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
DB 2139 TTATCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
QY 120 GAGTGCCTTCTTGACATCCCTGATCACTCTCAATCTTGTGGCTTTGAAGA 179
DB 2199 GAGTGCCTTCTTGACATCCCTGATCACTCTCAATCTTGTGGCTTTGAAGA 2258
QY 180 TACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCAGAGGTTCAAGGATAG 239
DB 2259 TACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCAGAGGTTCAAGGATAG 2318
QY 240 TCCCATCTATTTGGCGAGGATAGCCCAAGACTTGAAGCAATCTCATCTGAGAC 299
DB 2319 CCCCATCTATTTGGCGAGGATAGCCCAAGACTTGAAGCAATCTCATCTGAGAC 2378
QY 300 TTGTCTCTCGTAGGTGATGATTTACTTTTGGCCGCCATTCAGAAAAGCTTGCCATC 359
DB 2379 TTGTCTCTCGTAGGTGATGATTTACTTTTGGCCGCCATTCAGAAAAGCTTGCCATC 2438
QY 360 AAGGCAACCAAGGCTTTCATTTCTGCTCACTGTGGCTACATGTTTCCAAACCA 419
DB 2439 AAGGCAACCAAGGCTTTCATTTCTGCTCACTGTGGCTACATGTTTCCAAACCA 2498

QY 420 AGGCTCACTCTGCTCAAGAGGATTAAGGCTAAATTTATCCAAAGGACAGGGC 479
DB 2499 AGGCTCACTCTGCTCAAGAGGATTAAGGCTAAATTTATCCAAAGGACAGGGC 2558
QY 480 CCTCAGTGAAGAACATCCAG 501
DB 2559 CCTCAGTGAAGAACATCCAG 2580
RESULT 8
LOCUS AX000966 7582 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 11 from Patent WO9902696.
ACCESSION AX000966
VERSION AX000966.1 GI:7241208
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified
REFERENCE 1 (bases 1 to 7582)
AUTHORS Beseme, F. and Blond, J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 11 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)
FEATURES
source location/Qualifiers
1..7582 /organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others
ORIGIN
Query Match 88.2%; Score 442; DB 6; Length 7582;
Best Local Similarity 91.3%; Pred. No. 1e-120;
Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;
QY 1 CAAGATCTCAGATTAATCAAT-GAGGCTGTGTCCTCTATAGCCAGCTGTAAGCC 59
DB 2678 CAAGATCTCAGATTAATCAATGAGGCGTGTCTTTATACCAAGCTGTAAGCC 2727
QY 60 TTATCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
DB 2728 TTATCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2797
QY 120 GAGTGCCTTCTTGACATCCCTGATCACTCTCAATCTTGTGGCTTTGAAGA 179
DB 2798 GAGTGCCTTCTTGACATCCCTGATCACTCTCAATCTTGTGGCTTTGAAGA 2857
QY 180 TACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCAGAGGTTCAAGGATAG 239
DB 2858 TACTTCAAAACCAATCTCACTCACTGAGCTATTTTACCCAGAGGTTCAAGGATAG 2917
QY 240 TCCCATCTATTTGGCGAGGATAGCCCAAGACTTGAAGCAATCTCATCTGAGAC 298
DB 2918 YCCCATCTATTTGGCGAGGATAGCCCAAGACTTGAAGCAATCTCATCTGAGAC 2977
QY 299 -CTTGTCTCTCGTAGGTGATGATTTACTTTTGGCCGCCATTCAGAAAAGCTTGCC 357
DB 2978 TCTTGTCTCTCGTAGGTGATGATTTACTTTTGGCCGCCATTCAGAAAAGCTTGCC 3037
QY 358 TCAAGCAACCAAGGCTTTCATTTCTGCTCACTGTGGCTACATGTTTCCAAAC 417
DB 3038 TCAAGCAACCAAGGCTTTCATTTCTGCTCACTGTGGCTACATGTTTCCAAAC 3097
QY 418 AAGGCTCACTGCTCAAGAGG---TACTTAAAGGCTAAATTTATCCAAAGGAC 473
DB 3098 AARAGCTCACTGCTCAAGAGGTTAAATTAATTAATTAATTAATTAATTAATTA 3157
QY 474 CAGGAGCTTCAAGTGAAGAACATCCAG 501
DB 3158 CAGGAGCTTCAAGTGAAGAACATCCAG 3185

RESULT 9
AX027480 7582 bp DNA linear PAT 16-SEP-2000
LOCUS AX027480
DEFINITION Sequence 30 from Patent FR2788794.
ACCESSION AX027480
VERSION AX027480.1 GI:10188444
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 7582)
Mallet, F., Voissec, C. and Paranhos, B. G.
AUTHORS Mallet, F., Voissec, C. and Paranhos, B. G.
JOURNAL Patent: FR 2788794-A 30 28-JUL-2000;
BIO MERIEUX (FR)
FEATURES
source
1..7582
/location=Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others
ORIGIN
Query Match 88.2%; Score 442, DB 6; Length 7582;
Best Local Similarity 91.3%; Pred. No. 1e-120;
Matches 464; Conservative 31; Mismatches 6; Indels 7; Gaps 3;
QY 1 CAAGATCTCAGGATTATCAAT-GAGGCTGTGTCTCTATAGCCAGCTGTACTAGCCC 59
DB 2678 CAAGATCTCAGGATTATCAATGAGGCCGTTCTCTTTATACCAAGCTGTACTAGCCC 2737
QY 60 TTATACCTGCTTTTCCCAATACAGAGAGAGC/GAGTGTATTACAGTCTTGACCTTCA 119
DB 2738 TTATACCTGCTTTTCCCAATACAGAGAGAGC/GAGTGTATTACAGTCTTGACCTTMA 2797
QY 120 GATGCCCTTCTTCTGCATCCCTGTACATCTCTCAATCTTCTTGGTTGCTTGAAGA 179
DB 2798 GATGCCCTTCTTCTGCATCCCTGTACATCTCTCAATCTTCTTGGTTGCTTGAAGA 2857
QY 180 TACTTCAACCAACATCTCACTCACTGACATTTTATCCCAAGGGTTCAGGGATAG 239
DB 2858 TACTTCAACCAACATCTCACTCACTGACATTTTATCCCAAGGGTTCAGGGATAG 2917
QY 240 TCCCATCTATTGGCCAGGCTTAGCCCAAGCTTAGCCCAATCTCTCATCTTGACA- 298
DB 2918 YCCCATCTATTGGCCAGGCTTAGCCCAAGCTTAGCCCAATCTCTCATCTTGACA- 2977
QY 299 -CTTGTCTTGGTGTAGTGTAGTATTACTTTGGCCGCCCTTGAAGAACTTGGCA 357
DB 2978 TCTTGTCTTGGTGTAGTGTAGTATTACTTTGGCCGCCCTTGAAGAACTTGGCA 3037
QY 358 TCAAGCCACCAAGGCTTCTCAATTTCTCGCTACCTGTGCTACATGGTTTCCAAAC 417
DB 3038 TCAAGCCACCAAGGCTTCTCAATTTCTCGCTACCTGTGCTACATGGTTTCCAAAC 3097
QY 418 AAAGCTCAACTCTGCTCAGCAGAGT---TACTTAGGGCTAAATTTCCAAAGCAC 473
DB 3098 AARACTCACTCTGCTCAGCAGAGTAAATATCTAGGCTAARATTTCCAAAGGCGC 3157
QY 474 CAGGGCCCTCAGTGAAGAACATCCAG 501
DB 3158 CAGGGCCCTCAGTGAAGAAATTCAG 3185

REFERENCE
1 (bases 1 to 40205)
Taudien, S. and Rosenthal, A.
AUTHORS Taudien, S. and Rosenthal, A.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40205)
Taudien, S., Nordtsiek, G., Dagand, E., Hildmann, T., Drescher, B.,
AUTHORS Weber, J., Rosenthal, A. and Yaspo, M. L.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1998) Genome Analysis, Institut for Molecular
Biotechnology, Deutscherstrasse 11, Jena 07745, Germany
FEATURES
source
1..40205
/location=Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="cosmid Q11M15"
repeat_region complement(293..586)
/evidence=not experimental
/rpt_family="AluSc"
repeat_region complement(643..964)
/evidence=not experimental
/rpt_family="MERVL"
exon complement(1053..1176)
/note="Genscan, score = 5.19%, comment = Internal_exon 124
bp frame: 2 phase: 1"
/evidence=not experimental
/evidence=not experimental
/rpt_family="HERV16"
repeat_region complement(2296..3220)
/evidence=not experimental
/rpt_family="HERV16"
exon complement(2889..2889)
/note="Xpound exon prediction, score = 82% (0%)"
/evidence=not experimental
/evidence=not experimental
/evidence=not experimental
/rpt_family="MLT1C"
exon complement(3359..3395)
/note="Xpound exon prediction, score = 84% (0%)"
/evidence=not experimental
/evidence=not experimental
/evidence=not experimental
/rpt_family="L2"
exon complement(4385..4638)
/note="GRAIL, score = 95.000%, comment = excellent shadow"
/evidence=not experimental
/evidence=not experimental
/evidence=not experimental
/rpt_family="L1P1A14"
exon complement(6119..6598)
/note="GRAIL, score = 63.000%, comment = good"
/evidence=not experimental
/evidence=not experimental
/evidence=not experimental
/rpt_family="FLAM_A"
repeat_region complement(8048..8178)
/evidence=not experimental
/rpt_family="L2"
exon complement(8771..8876)
/note="MZEF, score = 67.4%"
/evidence=not experimental
/evidence=not experimental
/evidence=not experimental
/rpt_family="MER1A"
exon complement(8837..8876)
/note="GRAIL, score = 55.000%, comment = good shadow"
/evidence=not experimental
/evidence=not experimental
/evidence=not experimental
exon complement(9113..9213)

	Db	30326	CCTGTCCTTACGTGCATAGATGAATTACTTTTAGCTGCCCGCCTTGCAAAACCTTTGGCCAT	30267
Oy	359	CAAAGCACCCAAAGGCTCTTCAATTTCTCTCGTAACCTGTGGCTACAATGGTTTTCCAACCA	418	
Db	30266	CAAGCACCCAAAGGCTCTTAAAATTTCTCATACCTGTGGCTACAAAGGTTTCCAACCA	30207	
Oy	419	AAGGCTCAACTCTCTCTACAGCAGGTTA----	CTTAGGGCTAAATAATTCACAAAGCACC	474
Db	30206	AAGGCTCAGCTCTCTCTACAGCAGGTTAAATGCTTAGGGCTAAATAATTCACAAAGTCACC	30147	
Oy	475	AGGGCCCTCAGTGGAGAACAACATCCAG	501	
Db	30146	AGGGCCCTCAGTGGAGAACAACATTCACAG	30120	
RESULT	11			
LOCUS	AF121782/c	142742 bp	DNA	linear PRI 02-FEB-1999
DEFINITION	Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.			
ACCESSION	AF121782			
VERSION	AF121782.1	GI:4210991		
KEYWORDS	HTG.			
SOURCE	Homo sapiens.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 142742) Tad dien,S., Dagnid,E., Hildmann,T., Nordstiek,G., Drescher,B., Schatevov,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and Rosenthal,A. Direct Submission Submitted (21-JUN-1999) Genome Analysis, Institute for Molecular Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany Location/Dualifiers 1..142742 /organism="Homo sapiens;" /db_xref="taxon:9606" /chromosome="21" /map="21q22.3" /clone="PAC 206A10"			
TITLE	JOURNAL			
FEATURES	source			
	repeat_region	/evidence=not experimental		
	repeat_region	/rpt_family="MIR .415)		
	repeat_region	/evidence=not experimental		
	repeat_region	/rpt_family="MLRZE"		
	exon	/evidence=not experimental		
	repeat_region	/rpt_family="MER58A"		
	exon	/complement(1078..1162)		
	repeat_region	/note="GRAIL, score = 86.000%, comment = excellent shadow"		
	repeat_region	/evidence=not experimental		
	repeat_region	/rpt_family="MLTII"		
	repeat_region	/rpt_family="MLTII"		
	exon	/evidence=not experimental		
	repeat_region	/rpt_family="L2"		
	exon	/score="WZEP, score = 66.1%"		
	exon	/evidence=not experimental		
	repeat_region	/note="GRAIL, score = 99.000%, comment = excellent"		
	repeat_region	/evidence=not experimental		
	repeat_region	/complement(1683..1741)		
	repeat_region	/rpt_family="MIR"		
	repeat_region	/complement(1780..2326)		
	repeat_region	/evidence=not experimental		
	exon	/rpt_family="MLTIE"		
	exon	/complement(1834..1955)		
	exon	/note="GRAIL, score = 53.000%, comment = good shadow"		
	exon	/evidence=not experimental		
	repeat_region	/evidence=not experimental		
	repeat_region	/rpt_family="AlusX"		
	repeat_region	/rpt_family="AlusX"		
	repeat_region	/evidence=not experimental		
	repeat_region	/rpt_family="MLTIJ"		
	repeat_region	/evidence=not experimental		
	repeat_region	/rpt_family="THIB"		
	exon	/evidence=not experimental		
	exon	/rpt_family="MLTIJ"		
	exon	/complement(3502..3540)		
	exon	/note="GRAIL, score = 65.000%, comment = good shadow"		
	exon	/evidence=not experimental		
	exon	/note="GRAIL, score = 42.000%, comment = marginal"		
	exon	/evidence=not experimental		
	exon	/note="xround exon prediction, score = 70% (0%) "		
	exon	/evidence=not experimental		
	exon	/note="GRAIL, score = 95.1%"		
	exon	/evidence=not experimental		
	exon	/rpt_family="HBVL"		
	exon	/note="WZEP, score = 95.1%"		
	exon	/evidence=not experimental		
	exon	/note="GenScan, score = 1.83%, comment = initial_exon 6 bp frame: 1 phase: 0"		
	exon	/evidence=not experimental		
	exon	/note="GRAIL, score = 45.000%, comment = marginal"		
	exon	/evidence=not experimental		
	exon	/complement(8738..9232)		
	exon	/evidence=not experimental		
	exon	/rpt_family="MLTIC"		
	exon	/complement(9273..9420)		
	exon	/note="GRAIL, score = 71.000%, comment = good"		
	exon	/evidence=not experimental		
	exon	/note="GRAIL, score = 81.000%, comment = excellent shadow"		
	exon	/evidence=not experimental		
	exon	/note="homology = 100.00%, score = 38, counts = 2"		
	exon	/note="homology = 100.00%, score = 38, counts = 2"		
	exon	/rpt_type=tandem		
	exon	/rpt_unit=ttcttcgacactccctgtgagagcgcgttaccgc		
	exon	/evidence=not experimental		
	exon	/rpt_family="MLTIF"		
	exon	/evidence=not experimental		
	exon	/rpt_family="LMC/D"		
	exon	/evidence=not experimental		
	exon	/rpt_family="AluSG"		
	exon	/complement(11525..11777)		

exon /note="WZEF, score = 79.4%"
 /evidence=not_experimental
 complement(12346. .12827)
 repeat_region /evidence=not_experimental
 /rpt_family="MULTIC"
 12902. .13180
 /evidence=not_experimental
 /rpt_family="AluSg"
 complement(13057. .13161)
 exon /note="WZEF, score = 76.1%"
 /evidence=not_experimental
 complement(13326. .13364)
 repeat_region /evidence=not_experimental
 /rpt_family="L1MC2"
 complement(13613. .13908)
 repeat_region /evidence=not_experimental
 /rpt_family="AluSx"
 complement(14317. .14614)
 repeat_region /evidence=not_experimental
 /rpt_family="L1R37A"
 complement(14962. .15036)
 exon /note="WZEF, score = 93%"
 /evidence=not_experimental
 15255. .15299
 repeat_region /evidence=not_experimental
 /rpt_family="L1PA6"
 complement(15469. .15497)
 exon /note="Xpound exon prediction, score = 62% (0%)"
 repeat_region /evidence=not_experimental
 16045. .16768
 repeat_region /evidence=not_experimental
 /rpt_family="MER21A"
 complement(16555. .16655)
 exon /note="GRAL, score = 54.000%, comment = good"
 /evidence=not_experimental
 16823. .16850
 repeat_region /evidence=not_experimental
 /rpt_family="MER34"
 complement(16851. .17208)
 repeat_region /evidence=not_experimental
 /rpt_family="THE1A"
 17209. .17632
 repeat_region /evidence=not_experimental
 /rpt_family="MER34"
 complement(17633. .17969)
 repeat_region /evidence=not_experimental
 /rpt_family="AluSg1"
 complement(18004. .18334)
 repeat_region /evidence=not_experimental
 /rpt_family="L1R17"
 complement(18335. .18632)
 repeat_region /evidence=not_experimental
 /rpt_family="HERV17"
 18332. .18420
 exon /note="GRAL, score = 51.000%, comment = good shadow"
 /evidence=not_experimental
 complement(18397. .18584)
 exon /note="GRAL, score = 66.000%, comment = good"
 /evidence=not_experimental
 complement(18643. .24647)
 repeat_region /evidence=not_experimental
 /rpt_family="HERV17"
 complement(19228. .19361)
 exon /note="WZEF, score = 92.8%"
 /evidence=not_experimental
 20900. .21135
 exon /note="WZEF, score = 51.4%"
 /evidence=not_experimental
 21150. .21275
 exon /note="Xpound exon prediction, score = 74% (0%)"
 /evidence=not_experimental
 Query Match 86.8%; Score 435; DB 9; Length 142742;

Best Local Similarity 93.9%; Pred. No. 1e-118;
 Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
 QY 1 CAAGATCCAGAGATTATCAATGAGGCTGTGCTCTCATATGCGAGGCTGATACCTAGCCCT 60
 Db 20632 CAAGATCCAGAGATTATCAATGAGGCTGTGCTCTCATATGCGAGGCTGATACCTAGCCCT 20573
 QY 61 TATATCTGCTTTCCCAATATCCAGAGAGACAGATGTTTATAGTCTTGACCTTCCAG 120
 Db 20572 TATATCTGCTTTCCCAATATCCAGAGAGACAGATGTTTATAGTCTTGACCTTCCAG 20513
 QY 121 GATGCTTTCTTCTGATCCCTGTAACATCTGACCTCTCAATCTTGTGCTTTGAAAGAT 180
 Db 20512 GATGCTTTCTTCTGATCCCTGTAACATCTGACCTCTCAATCTTGTGCTTTGAAAGAT 20453
 QY 181 ACTTCAAAACCAATCTCAATCTCACTGACCTTTTATCCCAAGGCTTCAAGGATATG 240
 Db 20452 ACTTCAAAACCAATCTCAATCTCACTGACCTTTTATCCCAAGGCTTCAAGGATATG 20393
 QY 241 CCCCATCTATTGGCCAGGACCTTAGCCCAAGACTTGAGCCATCTCTCATACCTGACACT 300
 Db 20392 CCCCATCTATTGGCCAGGACCTTAGCCCAAGACTTGAGCCATCTCTCATACCTGACACT 20333
 QY 301 --TGCTTCGGTAGGTGATGATTTACTTTGGCCGCCATTCAGAAACCTTGCCAT 358
 Db 20332 CCTGCTTCGATGATGATGATTTACTTTAGCTGCCGCTGAGAAACCTTGCCAT 20273
 QY 359 CAAGCCACCCAGAGGCTCTTCAATTTCTGCTACCTGCTGCTACAGGCTTCCAAACCA 418
 Db 20272 CAAGCCACCCAGAGGCTCTTCAATTTCTGCTACCTGCTGCTACAGGCTTCCAAACCA 20213
 QY 419 AAGGCTCACTCTGCTCAGCAGAGTTA---CTTAGGGCTAAATTTATCCAAAGGACCC 474
 Db 20212 AAGGCTCACTCTGCTCAGCAGAGTTAATGCTTAGGGCTAAATTTATCCAAAGTACC 20153
 QY 475 AGGGCCTCAGTGAAGAACATTCAG 501
 Db 20152 AGGGCCTCAGTGAAGAACATTCAG 20126
 RESULT 12
 HS21C080/c 34000 bp DNA linear PRI 24-MAY-2000
 LOCUS
 DEFINITION Homo sapiens chromosome 21 segment HS21C080.
 ACCESSION AL163280 AP001735 BA000005
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 34000)
 REFERENCE
 AUTHORS
 Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T.,
 Park, H.S., Toyoda, A., Ishii, K., Toki, Y., Choi, D.K., Seoda, E.,
 Oki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K.,
 Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R.,
 Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A.,
 Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K.,
 Aakawa, S., Shintani, A., Sasaki, T., Nagamine, K., MitsuYama, S.,
 Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordström, G.,
 Honisch, K., Brandt, P., Scharfe, M., Scheen, O., Desario, A.,
 Reichelt, J., Kauer, G., Bioecker, H., Ramser, J., Beck, A., Klages, S.,
 Hennig, S., Rieseemann, L., Degand, E., Mehrtz, S., Borzym, K.,
 Gardiner, K., Nizetic, D., Francis, F., Leirich, H., Reinhardt, R. and
 Yaspo, M.L.
 TITLE
 JOURNAL
 Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
 Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
 Group * Institute of Molecular Biotechnology, Genome Analysis *
 Keio University School of Medicine, Dept. of Molecular Biology *
 GNF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
 Genetics (addresses see below)
 COMMENT
 The Chromosome 21 Mapping and Sequencing Consortium consists of


```

repeat_region      /rpt_type=DISPERSED
                    complement(11315. .11556)
                    /note="AluO"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
                    11702. .11753
                    /note="MIR"
                    /rpt_family="SINE/MIR"
                    /rpt_type=DISPERSED
                    12649. .12678
                    /gene="SH3BGR"
                    /number=5
                    complement(13115. .13425)
                    /note="AluSx"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
                    complement(13598. .13892)
                    /note="AluY"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
                    14013. .14131
                    /note="(CA)n"
                    /rpt_family="Simple_repeat"

```

Query Match	86.8%;	Score 435;	DB 9;	Length 340000;
Best Local Similarity	93.9%;	Pred. No. 1e-116;		
Matches 476;	Conservative	0;	Mismatches 25;	Indels 6;
				Gaps 2;

QY 1 CAAGATCTCAGGATTATCAATGAGGCTGTTGTTCCCTATAGCCAGCTGTACCTAGCCCT 60

Db 244792 CAAGATCTCAGGATTATCAATGAGGCTGTTTTCCTCTATACCACTATACCTAGCCCT 244733

61 TATACCTCTGCTTCCCAATACCAGAGGAGCAGAGTGGTTTACAGTCCCTGGACCTCAG 120

DB 244 / 32 TATACCTCGCTTCCCAATACGAGGAGCAGAGTGGTTACAGTCCGATCTTAAG 244673

[illegible][illegible][illegible]

09 341 CCCCCCTTTGGGCGCCAGCATTACCAATTAATTCGCGCAT 300

Db 244552 CCCCATTATTGGCCAGGCA TTAGCCCAAGACTTGA GCGCATACCTGGACACT 244493

301 --TGTCCTTCGGTAGGTGATGATTACTTTTGGCCGCCCATTCAGAAACCTTGTGCCAT 358

Db 244492 CCTGTCCTTCAGTGCATGGATGATTACTTTAGCTGCCCGTTCAGAAACCTTGTGCAT 244433

359 CAAGCCACCCAGCGCTCTCAATTCCTCGCTACCTGTGGCTACATGGTTCCAAACCA 418

Db 244432 CAAGCCACCCAGGCTCTTAATTTCCTCACTACCTGTGGCTACAAAGTTTCCAACCA 244373

419 AAGCTCAACTCTGCTCACAGCAGTTA---CTTAGGGCTAAAAATTATCCAAAGGCACC 474

DB 2443/2 AAGGCTCAGCTCTGCTCACGACAGTAAATGCTTAGGGCTAAATATATCCAAAGTCACC 244313

4/5 AGGGCCCCI CAG I GAGGAA CACAI CCAG 301

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

RESULT 13
81160000

LOCUS	AL162912	8339 bp	DNA	linear	PRI 22-NOV-20
DEFINITION	Human DNA sequence from clone PB6-198C4 on chromosome Xq13.3-31.3				

complete sequence.
ACCESSION AT162912 AT121824

VERSION ALL162912.1 GI: /406/22
KEYWORDS HTG.

SOURCE	ORGANISM
human.	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 8339)

JOURNAL COMMENT
Submitted (22-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: hummerj@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk During sequence assembly data is compiled from overlapping clones using differences and found that some overlapping clones

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/chrom/chx/>

THIS SEQUENCE WAS INHIBITED AS FOLLOWS UNLESS OTHERWISE NOTED: ALL REGIONS WERE EITHER DUBBED-STRANDED OR SEQUENCED WITH AN ALTERNATE CHEMISTRY OR COVERED BY HIGH QUALITY DATA (I.E., PHRED QUALITY >= 30). AN ATTEMPT WAS MADE TO RESOLVE ALL SEQUENCING PROBLEMS, SUCH AS COMPRESSIONS AND REPEATS: ALL REGIONS WERE COVERED BY AT LEAST ONE PLASMID SUBCLONE OR MORE THAN ONE M13 SUBCLONE; AND THE ASSEMBLY WAS CONFIRMED BY RESTRICTION DIGEST. RfE-198C4 IS FROM THE LIBRARY RfC1-6 CONSTRUCTED BY THE GROUP OF PIETER DE JONG. FOR FURTHER DETAILS SEE <http://www.chori.org/bacpac/home.htm>

VECTOR: pPAC4

IMPORTANT: THIS SEQUENCE IS NOT THE ENTIRE INSERT OF CLONE RfE-198C4 IT MAY BE SHORTER BECAUSE WE SEQUENCE OVERLAPPING SECTIONS ONLY ONCE, EXCEPT FOR A SHORT OVERLAP.

FEATURES	Location/Qualifiers
source	1. 8339
	/Location="Home" and "and"

```

/db_xref="taxon:9606"
/chromosome="X"
/map="q13.3-21.2"
/clone="RP6-198C4"
/clone_lib="RPCL-6"
repeat region
1..5641

```

```

repeat_region 5635, .5673 /NOBS=1100, REPEAT=100, MACR=100, CONDENS=100

```

repeat _{max} _region	.5955				
	5664.				

```
repeat_region      5956.  .6136      -
/note="ITP17 repeat: matches 1 310 of consensus"
```

```
repeat_region 6195. .6280
/notes="1:TR17 repeat: matches 239. 324 of consensus"
```

BASE COUNT	2649 a	1846 c	1726 g	2118 t
ORIGIN				

Query Match	85.5%;	Score 428.6;	DB 9;	Length 8339;
-------------	--------	--------------	-------	--------------

Matches 472; Conservative 0; Mismatches 29; Indels 6; Gaps 2;

1 CAAGATCTCAGGATTATCATGAGCGTGTGTTCCCTATAGCCAGCTGTACCTAGCCCT 60

Db 3668 CAAGATCTCAGGATTATCAGTGAGGCTGTTGTTTCTCTATACCCAGCTGTACCTAGCCCT 3727

61 TATACTCTGCTTCCCAAATACCAGAGGAGCAGAGTGGTTACAGTCCCTGGACCTTCAG 120

DB 3/28 1A1AC1C1G111CCCAA1ACAGAGGAAGCAGAG1GG11ACAG1CC1GGACG1AAAG 3/8/

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

|||||
Db 3788 GATGCTTTTCTGATCCCTGTAATCTGACTCTCAATCTTGTTGCCCTTGAGAT 3847
Qy 181 ACTTAAACCCAACTCTCACTGACCTGACCTATTTAAACCCAAAGGTTGAGGATGT 240
Db 3848 CTTTAAACCCAACTCTCACTGACCTGACCTATTTAAACCCAAAGGTTGAGGATGT 3907
Qy 241 CCCCATTATTTGGCGAGCATTTAGCCCAAGACTGAGCCCAATCCTCACTACCTGAGCA-- 298
Db 3908 CCCCATTATTTGGCGAGCATTTAGCCCAAGACTGAGCCCAATCCTCACTACCTGAGCACT 3967
Qy 299 CTTGCTCTTGGTAGGAGTATTTACTTTTG3CCGCCCATTCAGAAACCTTGCCAT 358
Db 3968 CTTGCTCTTGGTAGGAGTATTTACTTTTA3CCACCCATTCAGAAACCTTGCCAT 4027
Qy 359 CAAGCACCACCAAGCCTCTTCAATTTCTCGTACCTGCTAGTATGTTTCCAAACCA 418
Db 4028 CAAGCACCACCAAGCCTCTTCAATTTCTCGTACCTGCTAGTATGTTTCCAAACCA 4087
Qy 419 AAGGCTCAACTCTGCTCAGCAGAGT---TACTTAGGGCTAAATTTCCAAAGCACC 474
Db 4088 AAGGCTCAACTCTGCTCAGCAGAGTAACTTAAGGCTAAATTTCCAAAGCACC 4147
Qy 475 AAGGCTCAACTCTGCTCAGCAGAGTAACTTAAGGCTAAATTTCCAAAGCACC 501
Db 4148 AAGGCTCAACTCTGCTCAGCAGAGTAACTTAAGGCTAAATTTCCAAAGCACC 4174

RESULT 14

AX000964 3372 bp DNA linear PAT 10-MAR-2000
LOCUS Sequence 9 from Patent WO9902696.
DEFINITION AX000964
ACCESSION AX000964 GI:7241206
VERSION
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.

REFERENCE 1 (bases 1 to 3372)
AUTHORS Beeme,F. and Blond,J.
TITLE ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE
DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL Patent: WO 9902696-A 9 21-JAN-1999;
BIO MERIEUX (FR); BESEME FREDERIC (FR)

FEATURES
Location/Qualifiers
Source 1..3372
/organism="unidentified";
/db_xref="taxon:32644"
BASE COUNT 1047 a 835 c 711 g 779 t
ORIGIN

Query Match 85.5%; Score 428.4; DB 6; Length 3372;
Best Local Similarity 99.8%; Pred. No. 1.2e-116;
Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 72 TTCCCAATATACAGAGAGAGAGAGTGTTCAGTCTGACCTTCAGATGCTTCTT 131
Db 4 TTCCCAATATACAGAGAGAGAGAGTGTTCAGTCTGACCTTCAGATGCTTCTT 63
Qy 132 CTGATCCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAAGATATTTCAACC 191
Db 64 CTGATCCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAAGATATTTCAACC 123
Qy 192 AACATCTCACTACCTGAGCTATTTTACCCCAAGGTTCAAGGATAGTCCCATATTT 251
Db 124 AACATCTCACTACCTGAGCTATTTTACCCCAAGGTTCAAGGATAGTCCCATATTT 183
Qy 252 TGGCAGAGGATTAAGCCCAAGACTTGAAGCAATCTCATCTGAGACATTTGCTCGGT 311
Db 184 TGGCAGAGGATTAAGCCCAAGACTTGAAGCAATCTCATCTGAGACATTTGCTCGGT 243
Qy 312 AGGTGATATTTACTTTTGGCGCCCATTCAGAAACCTTGTCATCAAGCACCAG 371

Db 244 AGGTGATATTTACTTTTGGCGCCCATTCAGAAACCTTGTCATCAAGCACCAG 303
Qy 372 CGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAAGGCTCACTT 431
Db 304 CGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAAGGCTCACTT 363
Qy 432 GCTCAGAGGATTAAGGCTTAAATTTATCCAAAGGACCAAGGCTCCTGATGAGA 491
Db 364 GCTCAGAGGATTAAGGCTTAAATTTATCCAAAGGACCAAGGCTCCTGATGAGA 423
Qy 492 ACACATCCAG 501
Db 424 ACACATCCAG 433

RESULT 15

AX027478 3372 bp DNA linear PAT 16-SEP-2000
LOCUS AX027478
DEFINITION Sequence 28 from Patent FR2788784.
ACCESSION AX027478
VERSION AX027478.1 GI:10188442
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 3372)
AUTHORS Mallet,F., Voisset,C. and Paranhos,B.G.
JOURNAL Patent: FR 2788784-A 28 28-JUL-2000;
BIO MERIEUX (FR)

FEATURES
Location/Qualifiers
Source 1..3372
/organism="Homo sapiens";
/db_xref="taxon:9606"
BASE COUNT 1047 a 835 c 711 g 779 t
ORIGIN

Query Match 85.5%; Score 428.4; DB 6; Length 3372;
Best Local Similarity 99.8%; Pred. No. 1.2e-116;
Matches 429; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 72 TTCCCAATATACAGAGAGAGAGAGTGTTCAGTCTGACCTTCAGATGCTTCTT 131
Db 4 TTCCCAATATACAGAGAGAGAGAGTGTTCAGTCTGACCTTCAGATGCTTCTT 63
Qy 132 CTGATCCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAAGATATTTCAACC 191
Db 64 CTGATCCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAAGATATTTCAACC 123
Qy 192 AACATCTCACTACCTGAGCTATTTTACCCCAAGGTTCAAGGATAGTCCCATATTT 251
Db 124 AACATCTCACTACCTGAGCTATTTTACCCCAAGGTTCAAGGATAGTCCCATATTT 183
Qy 252 TGGCAGAGGATTAAGCCCAAGACTTGAAGCAATCTCATCTGAGACATTTGCTCGGT 311
Db 184 TGGCAGAGGATTAAGCCCAAGACTTGAAGCAATCTCATCTGAGACATTTGCTCGGT 243
Qy 312 AGGTGATATTTACTTTTGGCGCCCATTCAGAAACCTTGTCATCAAGCACCAG 371
Db 244 AGGTGATATTTACTTTTGGCGCCCATTCAGAAACCTTGTCATCAAGCACCAG 303
Qy 372 CGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAAGGCTCACTT 431
Db 304 CGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCAAGGCTCACTT 363
Qy 432 GCTCAGAGGATTAAGGCTTAAATTTATCCAAAGGACCAAGGCTCCTGATGAGA 491
Db 364 GCTCAGAGGATTAAGGCTTAAATTTATCCAAAGGACCAAGGCTCCTGATGAGA 423
Qy 492 ACACATCCAG 501
Db 424 ACACATCCAG 433

Mon Apr 21 09:59:17 2003

us-09-719-554-3_copy_5000_5500_1.rge

Page 15

Search completed: April 17, 2003, 06:41:04
Job time : 1888.47 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 17, 2003, 02:30:00 ; Search time 180.787 Seconds
(without alignments)
6240.778 Million cell updates/sec

Title: US-09-719-554-3_COPY_5000_5500
Perfect score: 501
Sequence: 1 caagatctcagatattacaa.....tcagtggagcaacatccag 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	501	100.0	10499	21	Human retroviral s
2	501	100.0	56093	24	Colion adenocarcino
3	474	94.6	2938	20	Human endogenous r
4	474	94.6	2938	21	Gag and partial po
5	474	94.6	7466	23	DNA encoding novel
6	453.8	90.6	3831	23	DNA encoding novel
7	442	88.2	7582	20	Complete human end
8	442	88.2	7582	21	Human endogenous r
9	435	86.8	3867	23	DNA encoding novel

10	435	86.8	3867	23	AA576464	DNA encoding novel
11	435	86.8	3867	23	AA580471	DNA encoding novel
12	435	86.8	4349	23	AA567475	DNA encoding novel
13	435	86.8	4535	23	AA576205	DNA encoding novel
14	435	86.8	5154	23	AA567609	DNA encoding novel
15	428.4	85.5	3372	20	AA525663	Human endogenous r
16	428.4	85.5	3372	21	AA59213	Partial pol gene a
17	422.2	84.3	900	23	AA567599	DNA encoding novel
18	422.2	84.3	900	23	AA572228	DNA encoding novel
19	422.2	84.3	900	23	AA573911	DNA encoding novel
20	422.2	84.3	900	23	AA576470	DNA encoding novel
21	422.2	84.3	1493	23	AA567606	DNA encoding novel
22	422.2	84.3	1687	23	AA572234	DNA encoding novel
23	422.2	84.3	1729	23	AA577018	DNA encoding novel
24	422.2	84.3	1743	23	AA584207	DNA encoding novel
25	422.2	84.3	2304	19	AA543199	Multiple sclerosis
26	420.6	84.0	924	23	AA572224	DNA encoding novel
27	420.6	84.0	1660	23	AA573608	DNA encoding novel
28	420.6	84.0	1660	23	AA576459	DNA encoding novel
29	420.6	84.0	1660	23	AA584201	DNA encoding novel
30	420.6	84.0	1810	23	AA585801	DNA encoding novel
31	420.6	84.0	1810	23	AA586023	DNA encoding novel
32	420.6	84.0	1846	23	AA573605	DNA encoding novel
33	420.6	84.0	1846	23	AA584192	DNA encoding novel
34	420.6	84.0	2220	23	AA567604	DNA encoding novel
35	420.6	84.0	2220	23	AA573919	DNA encoding novel
36	420.6	84.0	2220	23	AA576473	DNA encoding novel
37	419.6	83.8	1158	16	AA572815	Multiple sclerosis
38	419.6	83.8	1158	17	AA57973	Partial sequence o
39	419.6	83.8	1158	18	AA576527	MSRV-1B pol.* seque
40	419.6	83.8	1158	19	AA543115	Multiple sclerosis
41	419.6	83.8	2365	19	AA543200	Multiple sclerosis
42	419.6	83.8	2391	18	AA576478	MSRV-1 pol gene se
43	419.6	83.8	2391	19	AA543166	Multiple sclerosis
44	419	83.6	1117	23	AA576460	DNA encoding novel
45	419	83.6	1389	23	AA567594	DNA encoding novel

ALIGNMENTS

RESULT 1
ID ABN97929 standard; DNA; 10499 BP.
XX
AC ABN97929;
XX
DT 01-AUG-2002 (first entry)
XX
DE Human retroviral sequence HERV-7q.
XX
KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KW multiple sclerosis; ds.
XX
OS Human retrovirus.
XX
PN WO9967395-A1.
XX
PD 29-DEC-1999.
XX
PF 23-JUN-1999; 99WO-FR01513.
XX
PR 23-JUN-1998; 98FR-0007920.
XX
PA (INRM) INSEPM INST NAT SANTE & RECH MEDICALE.
XX
PI Alliel PM, Perin J, Rieger F;
XX
DR WPI; 2000-160587/14.
XX
FT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases -

XX Claim 3; Fig 1; 225bp; French.
XX
PS
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;

Query Match 100.0%; Score 501; DB 21; Length 10499;
Best Local Similarity 100.0%; Pred. No. 1.6e-151;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGATCTCAGGATTAATCAATGAGCTGTTGTTCTCTATAGCCAGCTGATACCTT 60
DB 5000 CAAGATCTCAGGATTAATCAATGAGCTGTTGTTCTCTATAGCCAGCTGATACCTT 5059
QY 61 TTTACTCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB 5060 TTTACTCTGCTTCCCAATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY 121 GATGCTCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 180
DB 5120 GATGCTCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCT 5179
QY 181 ACTTCAACCCCAATCTCACTCACTGAGCTATTATACCCCAAGGTTCCAGGATAGT 240
DB 5180 ACTTCAACCCCAATCTCACTCACTGAGCTATTATACCCCAAGGTTCCAGGATAGT 5239
QY 241 CCCCACTATTGTCGAGGATTAAGCCCAAGACTGAGCAATCCCATACCTGAGCACT 300
DB 5240 CCCCACTATTGTCGAGGATTAAGCCCAAGACTGAGCAATCCCATACCTGAGCACT 5299
QY 301 TGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 5300 TGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5359
QY 361 AGCCACCCCAAGGCTCTTCAATTTCTGCTTACCTGCTTACCTGCTTACCTGCTT 420
DB 5360 AGCCACCCCAAGGCTCTTCAATTTCTGCTTACCTGCTTACCTGCTTACCTGCTT 5419
QY 421 GGCTCAACTCTGCTCAACAGAGGTTACTTAAGGCTTAATTTCCAAAGACCCAGGCGC 480
DB 5420 GGCTCAACTCTGCTCAACAGAGGTTACTTAAGGCTTAATTTCCAAAGACCCAGGCGC 5479
QY 481 CTCAGTGAAGAACATCCAG 501
DB 5480 CTCAGTGAAGAACATCCAG 5500

RESULT 2
ABL61744
ID ABL61744 standard; DNA; 56093 BP.
XX
AC ABL61744;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilms' tumor; adenocarcinoma;

KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-235133P.
PR 18-SEP-2000; 2000US-235617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

(AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Sopet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
XX
XX Claim 1; SEQ ID 81; 44bp; English.
XX
XX The present invention describes a method (M1) for screening for an
XX anti-neoplastic agent. The method involves exposing cells to a chemical
XX agent to be tested for anti-neoplastic activity, determining a change in
XX expression of at least one gene (I) of a signature gene set, where (i)
XX comprises a sequence (S) selected from 8447 sequences (given in ABL6164
XX to ABL70110), or is at least 95% identical to (S), where a change in
XX expression is indicative of anti-neoplastic activity. (i) has cytostatic

AC AAS9206;
XX 07-NOV-2000 (first entry)
DE Gag and partial pol gene fragment of HERV-W from human genome.
XX
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
OS Homo sapiens.
XX MO200043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000MO-FR00144.
XX PF
XX 21-JAN-1999; 99FR-0000889.
XX
XX (INMR) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy,
PT contains at least part of the gag gene -
XX
XX
XX Disclosure; Page 43; 53pp; French.
XX
XX The present sequence represents an endogenous retroviral nucleic acid
CC fragment, which is associated with an autoimmune disease, and is
CC integrated into the human genome. The fragment is originally derived
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
CC HERV-W retrovirus is associated with autoimmune disease, failure of
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or
CC proteins derived from it, are useful for diagnosis of autoimmune
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
CC The nucleic acid fragments may also be used for in situ labelling of
CC isolated chromosomes, while the transcription product can be used to
CC study or monitor T cell proliferation in vitro.
XX
XX Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;
XX
XX Query Match 94.6%; Score 474; DB 21; Length 2938;
XX Best Local Similarity 97.8%; Pred. No. 5e-143;
XX Matches 491; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
XX
QY 1 CAGATCTCAGGATTTATCAAT-GAGGCTGTTGTCTCTATAGCAGCTGATCTAGGCC 59
DB 2079 CAGATCTCAGGATTTATCAATGAGGCGCTTGTCTTTATACCAAGCTGATCTAGGCC 2138
QY 60 TTATACCTGCTTTTCCCAATACAGAGAGACAGAGGTTTAAAGTCTGACCTTCA 119
DB 2139 TTATACCTGCTTTTCCCAATACAGAGAGACAGAGGTTTAAAGTCTGACCTTCA 2198
QY 120 GGATGCTTCTCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTG 179
DB 2199 GGATGCTTCTCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTGATCCCTG 2258
QY 180 TACTTCAAAACCAAGATCTCACTCACTGACCTGATCTTATACCCCAAGGTTGAGGATAG 239
DB 2259 TACTTCAAAACCAAGATCTCACTCACTGACCTGACCTGATCTTATACCCCAAGGTTG 2318
QY 240 TCCCATCTATTGTCGACGATAGCCCAAGATTGAGCCATCTCTATCTGACAC 299
DB 2319 CCCCATCTATTGTCGACGATAGCCCAAGATTGAGCCATCTCTATCTGACAC 2378
QY 300 TTGTCCTTGGAGAGGTGATGATTTAATTTTGGCGCCCATTCAGAAACCTTGTGCATC 359
DB 2379 TTGTCCTTGGAGAGGTGATGATTTAATTTTGGCGCCCATTCAGAAACCTTGTGCATC 2438

QY 360 AAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 419
DB 2439 AAGCCACCAAGGCTCTTCAATTTCTGCTACCTGTGCTACATGTTTCCAAACCA 2498
QY 420 AGGTCACCTCTGCTCAGAGGTTACTTAGGCTAAATATCCAAAGCACACAGGC 479
DB 2499 AGGTCACCTCTGCTCAGAGGTTACTTAGGCTAAATATCCAAAGCACACAGGC 2558
QY 480 CCTCAGTGGAGACACATCCAG 501
DB 2559 CCTCAGTGGAGACACATCCAG 2580
RESULT 5
ID AAS68626 standard; cDNA; 7466 BP.
XX
XX AAS68626;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #4430.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG04439.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 4430; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy involving
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

RESULT 7
 AAX25665
 ID AAX25665 standard; cDNA to mRNA; 7582 BP.
 AC AAX25665;
 XX
 XX
 XX 21-MAY-1999 (first entry)
 XX
 XX Complete human endogenous retrovirus W genome.
 XX
 XX Clone: human endogenous retrovirus; genome; autoimmune disease;
 KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 XX
 XX Human endogenous retrovirus.
 OS
 XX MO9902696-A1.
 PN
 XX 21-JAN-1999.
 PD
 XX 06-JUL-1998; 98WO-FR01442.
 PF
 XX 07-JUL-1997; 97FR-0008815.
 PR
 XX (IMMR) BIO MERIEUX.
 PA
 XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
 PI WPI; 1999-120897/10.
 DR
 XX New nucleic acid sequences from human endogenous retrovirus-W -
 PT expressed exclusively in placenta and useful in diagnosis and
 PT therapy of autoimmune disease, and abnormal or failed pregnancy
 XX
 XX Claim 1; Page 71-74; 106pp; French.
 PS
 XX This sequence represents the complete sequence of the human endogenous
 CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
 CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
 CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
 CC insulin-dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility.
 CC
 CC Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;
 SQ
 Query Match 88.2%; Score 442; DB 20; Length 7582;
 Best Local Similarity 91.3%; Pred. No. 1.8e-132;
 Matches 464; Conservative 31; Mismatch 6; Indels 7; Gaps 3;

Db 2978 TCTTGCTTCRGAKGTGATGATTACTTTTTCGCGCCYRTTCAGAAACCTTGCCA 3037
 Qy 358 TCAAGCCACCAGACGGCTTTCATTTCTCGTACTGTGGCTACATGGTTCCAAACC 417
 Db 3038 TCAAGCCACCAGACGCTTTCATTTCTCGTACTGTGGCTACATGGTTCCAAACS 3097
 Qy 418 AAAGCTCACTCTGCTCCACAGCAGT----TACTTGGGCTAAATATCCAAAGGCAC 473
 Db 3098 ARAAGCTCACTCTGCTCCACAGCAGTAAATACTTAGGRCCTAATATTCCAAAGGCAC 3157
 Qy 474 CAGGCGCTCAGTGTGAGGAGACATCCAG 501
 Db 3158 CAGGCGCTCAGTGTGAGGAGATATTCAG 3185
 RESULT 8
 ID AAA59215 standard; DNA; 7582 BP.
 AC AAA59215;
 XX
 XX 07-NOV-2000 (first entry)
 DT
 XX Human endogenous retrovirus W (HERV-W) sequence.
 DE
 XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
 KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
 KM
 XX Human endogenous retrovirus.
 OS
 XX
 FH Key Location/Qualifiers
 FT LTR 1..120
 FT /*tag= a
 FT /note= "R of 5' LTR"
 FT 121..575
 FT /*tag= b
 FT /note= "U5 of 5' LTR"
 FT 579..596
 FT /*tag= c
 FT CDS 5581..7194
 FT /*tag= d
 FT /note= "ORF1 env538"
 FT 7039..7194
 FT /*tag= e
 FT /note= "ORF2 52 AA"
 FT 7112..7255
 FT /*tag= f
 FT /note= "ORF3 48 AA"
 FT 7244..7254
 FT misc_feature
 FT /*tag= g
 FT /note= "polypurine tract"
 FT 7256..7582
 FT /*tag= h
 FT /note= "U3-R of 3' LTR"
 FT 7563..7569
 FT /*tag= i
 FT
 XX MO200043521-A2.
 PN
 XX 27-JUL-2000.
 PD
 XX 21-JAN-2000; 2000WO-FR00144.
 PF
 XX 21-JAN-1999; 99FR-0000888.
 PR
 XX (IMMR) BIO MERIEUX.
 PA
 XX Paranhos-Baccala G, Mallet F, Voisset C;
 PI WPI; 2000-499229/44.
 DR
 XX New nucleic acid from human endogenous retrovirus, useful e.g. for
 PT diagnosis of autoimmune disease and complications of pregnancy.

Db 1993 CCTGCTCCTCAGTGCATGATGATTACTTTTA3CTGCCCGCTTCAGAAACCTTGGCCAT 2052
Qy 359 CAAGCCACCCAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGATTTCCAAACCA 418
Db 2053 CAAGCCACCCAGCGCTCTTAAATTTCTCCTCACTACCTGTGGCTACAGATTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGTCTCAGCAGAGGTTA----CTTAGGGCTTAAATTTTCCAAAGGCAAC 474
Db 2113 AAGGCTCAGCTCTGCTCTCAGCAGGTTAAATGCTTAGGGCTTAAATTTTCCAAAGTCACC 2172
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 2173 AGGGCCCTCAGTGAAGAACGTAATCCAG 2199

RESULT 10
AAS76464 standard; cDNA; 3867 BP.
Xx AAS76464;
Ac AAS76464;
Dt 13-FEB-2002 (first entry)
Xx DNA encoding novel human diagnostic protein #12268.
Xx Human; chromosome mapping; gene mapping; gene therapy; forensic;
Kw food supplement; medical imaging; diagnostic; genetic disorder; ss.
Xx Homo sapiens.
Xx WO200175067-A2.
Xx 11-OCT-2001.
Xx 30-MAR-2001; 2001WO-US08631.
Xx 31-MAR-2000; 2000US-0540217.
Pr 23-AUG-2000; 2000US-0649167.
Xx (HYSE-) HYSEQ INC.
Xx Drmanac RT, Liu C, Tang YT;
Xx WPI; 2001-639362/73.
Dr P-PSDB; ABG12277.
Xx New isolated polynucleotide and encoded polypeptides, useful in
Pt diagnostics, forensics, gene mapping, identification of mutations
Pt responsible for genetic disorders or other traits and to assess
Pt biodiversity -
Xx Claim 1; SEQ ID No 12268; 103bp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridization probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, its molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO

CC at fdp.wipo.int/pub/published_pct_sequences.
Xx Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
SQ
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best local Similarity 93.9%; Pred. No. 2,5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
Qy 1 CAAGATCTCAGGATTAATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCT 60
Db 1693 CAAGATCTCAGGATTAATCAATGAGGCTGTGTTCTCTATAGCCAGCTGTACTAGCCCT 1752
Qy 61 TATCTCTGCTTTCCCAATACCAAGAGAGAGAGGTTTACAGTCTGAGACCTTCAG 120
Db 1753 TATCTCTGCTTTCCCAATACCAAGAGAGAGAGGTTTACAGTCTGAGACCTTCAG 1812
Qy 121 GATGCTCTTCTTGTCATCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAGAT 180
Db 1813 GATGCTCTTCTTGTCATCCTGTACATCTGACTCTCAATCTTGTGCTTTGAAGAT 1872
Qy 181 ACTTCAAAACCAACATCTCAATCTGACCTGACATTTTACCCCAAGGTTCCAGGATAGT 240
Db 1873 CTTCAAAACCAACATCTCAATCTGACCTGACATTTTACCCCAAGGTTCCAGGATAGT 1932
Qy 241 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCCAATCTCATACCTGGACACT 300
Db 1933 CCCCATCTATTTGGCCAGGATTTAGCCCAAGACTTGAGCCCAATCTCATACCTGGACACT 1992
Qy 301 --TGCTCTGGGTAGTGATGATTTACTTTTGGCCGCCCATTCAGAAACCTTGGCCAT 358
Db 1993 CCTGCTCCTCAGTGCATGATGATTTACTTTTGGCTGCTGCTGCTGCTGCTGCTGCT 2052
Qy 359 CAAGCCACCCAGCGCTCTTCAATTTCTCGTACCTGTGCTACATGATTTCCAAACCA 418
Db 2053 CAAGCCACCCAGCGCTCTTAAATTTCTCCTCACTACCTGTGGCTCAAGGTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGTCTCAGCAGGTTA----CTTAGGGCTTAAATTTTCCAAAGCACC 474
Db 2113 AAGGCTCAGCTCTGCTCTCAGCAGGTTAAATGCTTAGGGCTTAAATTTTCCAAAGTCACC 2172
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 2173 AGGGCCCTCAGTGAAGAACGTAATCCAG 2199

RESULT 11
AAS80471 standard; cDNA; 3867 BP.
Xx AAS80471;
Ac AAS80471;
Dt 13-FEB-2002 (first entry)
Xx DNA encoding novel human diagnostic protein #16275.
Xx Human; chromosome mapping; gene mapping; gene therapy; forensic;
Kw food supplement; medical imaging; diagnostic; genetic disorder; ss.
Xx Homo sapiens.
Xx WO200175067-A2.
Xx 11-OCT-2001.
Xx 30-MAR-2001; 2001WO-US08631.
Xx 31-MAR-2000; 2000US-0540217.
Pr 23-AUG-2000; 2000US-0649167.
Xx (HYSE-) HYSEQ INC.
Xx Drmanac RT, Liu C, Tang YT;
Xx

DR WPI, 2001-639362/73.
DR P-PSDB; ABG16284.
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 1; SEQ ID No 16275; 103bp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 3867 BP; 1148 A; 992 C; 838 G; 889 T; 0 other;
Query Match 86.8%; Score 435; DB 23; Length 3867;
Best Local Similarity 93.9%; Pred. No. 2.5e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGATCCTAGCCCT 60
DB 1693 CAAAGTCTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGATCCTAGCCCT 1752
QY 61 TATACTGTGTTCCCAATATCAGAGAGACAGAGTGGTTTACAGTCTGAGACCTTCAG 120
DB 1753 TATACTGTGTTCCCAATATCAGAGAGACAGAGTGGTTTACAGTCTGAGACCTTCAG 1812
QY 121 GATGCTCTTCTTGATCCTCTGATCCTCTGATCCTCAATCTTGTGTTGCTTGAAGAT 180
DB 1813 GATGCTCTTCTTGATCCTCTGATCCTCTGATCCTCAATCTTGTGTTGATGAAGAT 1872
QY 181 ACTTCAACCCCAATCTCACTCAGCTGACCTGACTATTTTAAACCCCAAGGTTGAGGATAGT 240
DB 1873 CTTTCAACCCCAATCTCACTCAGCTGACCTGACTATTTTAAACCCCAAGGTTGAGGATAGT 1932
QY 241 CCCCATCTATTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCCTTATCCTGAGACT 300
DB 1933 CCCCATCTATTGGCCAGGATTTAGCCCAAGACTTGAAGCAATCCTTATCCTGAGACT 1992
QY 301 --TGTCTTCCGTAAGTGAATGATTTACTTTTGGCCGCCATTCAGAAACCTTGTGCCAT 358
DB 1993 CTTGTCTTCCGTAAGTGAATGATTTACTTTTGGCCGCCATTCAGAAACCTTGTGCCAT 2052
QY 359 CAAGCCACCCCAAGGCTTTCATTTCTCGCTACCTGTGCTACATGAGTTTCCAAACCA 418
DB 2053 CAAGCCACCCCAAGGCTTTCATTTCTCGCTACCTGTGCTACATGAGTTTCCAAACCA 2112
QY 419 AAGGCTCAACTCTGCTCAGCAGAGTTA----CTTAGGGCTAAATTTTCCAAAGGCACC 474
DB 2113 AAGGCTCAACTCTGCTCAGCAGAGTTAATGCTTAGGGCTAAATTTTCCAAAGTCAAC 2172
QY 475 AGGGCCCTCAAGTGAAGAACACATCCAG 501
DB 2173 AGGGCCCTCAAGTGAAGAACATCCAG 2199

RESULT 12
AAS76475
ID AAS76475 standard; cDNA; 4349 BP.
XX
XX AAS76475;
AC
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #12279.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
XX WO200175067-A2.
PN
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
PI
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI, 2001-639362/73.
DR P-PSDB; ABG12288.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PS Claim 1; SEQ ID No 12279; 103bp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 4349 BP; 1319 A; 1061 C; 941 G; 1026 T; 2 other;
Query Match 86.8%; Score 435; DB 23; Length 4349;
Best Local Similarity 93.9%; Pred. No. 2.6e-130;
Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;
QY 1 CAAGATCTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGATCCTAGCCCT 60
DB 2382 CAAGATCTCAGATTATCAATGAGGCTGTGTCCTCTATAGCCAGCTGATCCTAGCCCT 2441
QY 61 TATACTGTGTTCCCAATATCAGAGAGACAGAGTGGTTTACAGTCTGAGACTTCAG 120

Db 2442 TATACCTGCTTCCCAATACCCAGAGAACCAAGTGTGTTAAGTCTGATCTTAAG 2501
Qy 121 GATGCTCTCTGTCGATCCCGTACATCTGACCTCAATCTGTTGGCTTTGAAGAT 180
Db 2502 GATGCTCTCTGTCGATCCCGTACATCTGACCTCAATCTGTTGGCTTTGAAGAT 2561
Qy 181 ACTTCAACCCCAACATCTCAACTGACCTGACTATTTTACCCTCAAGGTTGAGGATAGT 240
Db 2562 CCTTCAACCCCAACATCTCAACTGACCTGACTATTTTACCCTCAAGGTTGAGGATAGT 2621
Qy 241 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAGCCAACTCTATCTGAGACCT 300
Db 2622 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAGCCAACTCTATCTGAGACCT 2681
Qy 301 --TGTCCTTGGTGGTGGATGATTTACTTTTGACCGCCCATTTGAGAACTTGCCAT 358
Db 2682 CCGTCTCTTCAAGTGCATGATGATTTACTTTTACCTGCCCCCTTCAGAAACCTTGTCAT 2741
Qy 359 CAAGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 418
Db 2742 CAAGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 2801
Qy 419 AAGGCTCAACTCTGCTCAACAGAGGTTA---CTTAGGGCTAAATTTTCCAAAGGCACC 474
Db 2802 AAGGCTCAACTCTGCTCAACAGAGGTTAATGCTTAGGGCTAAATTTTCCAAAGTCACC 2861
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 2862 AGGGCCCTCAGTGAAGAACATCCAG 2888

RESULT 13
AAS76205
ID AAS76205 standard; cDNA, 4535 BP.
XX AAS76205;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #12009.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI, 2001-639362/73.
XX P-Psdb; ABG12018.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostic, forensic, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 12009; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4535 BP; 1356 A; 1159 C; 983 G; 1037 T; 0 other;

Query Match 86.8%; Score 435; DB 23; Length 4535;

Best Local Similarity 93.9%; Pred. No. 2, 7e-130;

Matches 476; Conservative 0; Mismatches 25; Indels 6; Gaps 2;

Qy 1 CAAGATTCAGGATTCATCAATGAGCTGTGTTGTTCTTATAGCACTGTACCTGACCT 60
Db 1693 CAAGATTCAGGATTCATCAATGAGCTGTGTTGTTCTTATAGCACTGTACCTGACCT 1752
Qy 61 TATACCTGCTTCCCAATACCCAGAGAGAGAGTGTTCACGCTGACCTTCAAG 120
Db 1753 TATACCTGCTTCCCAATACCCAGAGAGAGAGTGTTCACGCTGACCTTCAAG 1812
Qy 121 GATGCTCTCTGTCGATCCCGTACATCTGACCTCAATCTGTTGGCTTTGAAGAT 180
Db 1813 GATGCTCTCTGTCGATCCCGTACATCTGACCTCAATCTGTTGGCTTTGAAGAT 1872
Qy 181 ACTTCAACCCCAACATCTCAACTGACCTGACTATTTTACCCTCAAGGTTGAGGATAGT 240
Db 1873 ACTTCAACCCCAACATCTCAACTGACCTGACTATTTTACCCTCAAGGTTGAGGATAGT 1932
Qy 241 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAGCCAACTCTATCTGAGACCT 300
Db 1933 CCCCATCTATTTGGCCAGGACATTAAGCCCAAGCTTGAGCCAACTCTATCTGAGACCT 1992
Qy 301 --TGTCCTTGGTGGTGGATGATTTACTTTTGCCGCCCAATTCAGAAACCTTGTCAT 358
Db 1993 CCGTCTCTTCAAGTGCATGATGATTTACTTTTACTGCCCCGTTCAAGAACTTGTCAT 2052
Qy 359 CAAGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 418
Db 2053 CAAGCCACCCCAAGCGCTCTTCAATTTCTGCTACCTGTGGCTACATGTTTCCAAACCA 2112
Qy 419 AAGGCTCAACTCTGCTCAACAGCAGGTTA---CTTAGGGCTAAATTTTCCAAAGGCACC 474
Db 2113 AAGGCTCAACTCTGCTCAACAGCAGGTTAATGCTTAGGGCTAAATTTTCCAAAGTCACC 2172
Qy 475 AGGGCCCTCAGTGAAGAACATCCAG 501
Db 2173 AGGGCCCTCAGTGAAGAACATCCAG 2199

RESULT 14

AAS67609
ID AAS67609 standard; cDNA, 5154 BP.

XX AAS67609;
AC

XX 13-FEB-2002 (first entry)
DT

XX DNA encoding novel human diagnostic protein #3413.
XX

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW

QY 132 CTGATCCCTGTACATCCTGACTCTCAATTCGTTTGCTTTGAAGATACCTTCAAAACC 191
|||
Db 64 CTGATCCCTGTACATCCTGACTCTCAATTCGTTTGCTTTGAAGATACCTTCAAAACC 123
|||
QY 192 AACATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGTCCCATCTATT 251
|||
Db 124 AGCATCTCACTCACTGAGACTATTTTACCCCAAGGTTCAAGGATAGTCCCATCTATT 183
|||
QY 252 TGGCAGGCAATTAGCCCAAGACTGAGCCAAATCTCATACCTGGACACTTGTCCCTCGGT 311
|||
Db 184 TGGCAGGCAATTAGCCCAAGACTGAGCCAAATCTCATACCTGGACACTTGTCCCTCGGT 243
|||
QY 312 AGTGGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCCATCAGCCACCCAAAG 371
|||
Db 244 AGTGGATGATTTACTTTGGCCGCCCATTCAGAACTTGTGCCATCAGCCACCCAAAG 303
|||
QY 372 CGCTCTTCAATTTCTGCTACCTGTGCTACATGCTTTCCAAACCAAGGCTCACTCT 431
|||
Db 304 CGCTCTTCAATTTCTGCTACCTGTGCTACATGCTTTCCAAACCAAGGCTCACTCT 363
|||
QY 432 GCTCACAGCAGTTACTTAGGGCTTAAATTATCCAAAGCACAGGGGCTCAAGTAGAGA 491
|||
Db 364 GCTCACAGCAGTTACTTAGGGCTTAAATTATCCAAAGCACAGGGGCTCAAGTAGAGA 423
|||
QY 492 ACACATCCAG 501
|||
Db 424 ACACATCCAG 433
|||

Search completed: April 17, 2003, 05:07:04
Job time : 209.787 secs